Journal of Advanced Computing Systems (JACS)
www.scipublication.com

Integrating Machine Learning and Human Feedback for Employee Performance Evaluation

Kiran Kumar Reddy Yanamala
Central Michigan University

DOI: 10.69987/JACS.2022.20101

Keywords

Employee Performance Evaluation, Machine Learning (ML), Random Forest Algorithm, Data-Driven Feedback, Human Feedback, Performance Metrics.

Abstract

Employee performance evaluation has traditionally relied on subjective methods such as annual reviews, peer assessments, and self-reports, often leading to inconsistencies and biases. This study investigates the application of machine learning (ML), specifically the random forest algorithm, to enhance objectivity and consistency in performance evaluation. By leveraging large datasets, ML models offer data-driven feedback on key performance metrics, such as productivity and task completion rates, reducing the inherent subjectivity found in traditional systems. This research compares ML-driven feedback with human feedback across various performance categories, highlighting the advantages and limitations of both approaches. The findings suggest that while ML feedback offers superior consistency for quantifiable metrics, human feedback is essential for assessing soft skills such as leadership potential and emotional intelligence. The study also examines employee acceptance of ML feedback, identifying demographic factors that influence feedback receptivity. By integrating both ML-driven and human feedback systems, organizations can achieve more balanced, accurate, and fair performance evaluations. This research provides valuable insights into the future of performance management, where AI and human oversight work together to foster continuous development and transparency.

Introduction

Context of Traditional Performance Evaluation

Employee performance evaluation has traditionally relied on methods such as annual reviews, peer assessments, and self-reports. These systems typically aim to provide feedback on an employee's strengths and areas for improvement, guiding career development and organizational decisions. However, traditional methods are often plagued by significant limitations. One of the major criticisms of these approaches is the subjectivity involved, which can lead to personal biases in evaluations. For instance, Zbranek (2013) highlights the inefficiency of traditional methods, which often fail to provide objective and consistent assessments of employee performance [1]. Another limitation is the failure of traditional systems to provide real-time or continuous feedback. Most traditional evaluations occur annually or semi-annually, leading to feedback that is outdated or no longer relevant. Payne et al. (2009) compared online and traditional performance appraisal systems and found that employees reported higher satisfaction with systems that allowed for more frequent feedback and greater participation in the evaluation process [2]. Moreover, Bazerman et al. (1982) discussed the escalation of commitment in traditional evaluation systems, wherein raters who have previously promoted an employee may feel compelled to continue rating them positively, regardless of performance, thus skewing evaluations and hindering fairness [3]. Overall, traditional performance evaluation systems lack the necessary objectivity, timeliness, and fairness to effectively evaluate employees in today's fast-paced business environments. This has led to an increased demand for more advanced, data-driven approaches to performance evaluation.

Advances in AI and ML for Performance Evaluation

In response to the limitations of traditional evaluation systems, artificial intelligence (AI) and machine learning (ML) have emerged as transformative tools in performance management. These technologies offer a means to automate and optimize performance evaluations, reducing the inherent subjectivity of human assessments [4]–[7]. AI and ML models are particularly valuable because they can analyze large datasets in real time, allowing organizations to provide continuous feedback to employees [8], [9]. By integrating data from multiple sources, including task completion rates, peer feedback, and manager evaluations, these models can provide a holistic view of employee performance. Kacmar et al. (2009) examined the role of perceived work environment in traditional evaluations and emphasized that AI models could enhance this by integrating environmental factors into performance predictions, making the evaluations more accurate and less prone to human error [10]. Additionally, AI-based performance evaluation systems reduce biases that are often present in human-led assessments [11]. As highlighted by Reich and Barai (1999), machine learning models have the ability to process data in a way that removes the cognitive biases that can affect human raters [12]. Another critical advancement in AI for performance evaluation is the application of fuzzy logic systems. Manoharan et al. (2011) introduced a fuzzy multi-attribute decision-making model to performance evaluation, demonstrating that fuzzy logic can better manage the complexities of employee performance appraisal by focusing on various factors simultaneously and reducing information loss, which is common in traditional appraisal systems [13].

This study makes several important contributions to the field of employee performance evaluation by exploring the integration of machine learning (ML) and traditional human feedback systems. First, it demonstrates how ML-driven feedback can improve the objectivity and consistency of evaluations by reducing the subjectivity and biases inherent in traditional methods, particularly for quantifiable metrics like productivity and task completion rates. The research highlights the complementary strengths of ML and human feedback systems, proposing a hybrid approach that combines objective performance data with qualitative insights, such as leadership potential and emotional intelligence, provided by human evaluators. Additionally, the study examines the variability in employee acceptance of ML feedback across different demographics, offering valuable insights that can help organizations tailor feedback mechanisms to enhance engagement and acceptance. Through the application of advanced ML techniques, such as random forests and fuzzy logic systems, the research provides a practical framework for implementing ML-based performance evaluations, with a focus on reducing bias and improving fairness. Furthermore, the study provides guidelines for integrating ML and human feedback, recommending a balanced model where ML evaluates task-related performance, while human feedback addresses more nuanced, interpersonal aspects of performance. Overall, this study contributes to the growing use of AI and ML in performance management, offering solutions to enhance the effectiveness, fairness, and transparency of employee evaluations.

MACHINE LEARNING AND DATA ANALYTICS IN PERFORMANCE EVALUATION

Overview of Random Forest Algorithm in Employee Performance Evaluation

In this study, we employ the Random Forest algorithm, proposed by Breiman (2001), to predict employee performance [14]–[17]. This ensemble learning method operates by constructing multiple decision trees using various features of employee data, such as task completion rate, peer feedback, manager evaluations, attendance records, and other performance-related metrics. Each tree in the Random Forest predicts employee performance based on a subset of the data, and the final prediction is obtained by aggregating the individual tree outputs.

Mathematically, the forest consists of *M* decision trees, each built using a bootstrapped sample from the original dataset. For each tree *j*, the prediction at a given input *x* (which corresponds to the features of an employee) is represented as $m_n(x;\Theta_i, D_n)$, where Θ_i represents the random variables that control both the bootstrap sampling and feature selection for that tree. The final forest prediction is obtained by averaging the predictions of all trees:

$$m_{M,n}(x;\Theta_1,\ldots,\Theta_M, D_n) = \frac{1}{M}\sum_{j=1}^{M} m_n(x;\Theta_j, D_n) \quad (1)$$

In our employee performance prediction model, the x-values represent the performance metrics of an employee, such as task completion, peer feedback, and attendance records, and the response variable Y represents the predicted performance level (e.g., high performer, medium performer, or low performer). This aggregation over multiple trees reduces variance and helps mitigate the risk of overfitting, making the model robust even in high-dimensional feature spaces.

Algorithmic Implementation for Employee Performance Prediction

In the context of employee performance evaluation, the Random Forest model constructs each decision tree by recursively partitioning the employee feature space. For each node in the tree, a random subset of features (e.g., task completion rate, peer feedback, etc.) is selected from the full feature set, and the best feature-split is determined using the CART-split criterion. The criterion for regression is based on minimizing the mean squared error (MSE), while for classification, it focuses on maximizing the reduction in Gini impurity or entropy. For example, in predicting an employee's performance score *Y*, the CART-split criterion at a node is mathematically defined as:

$$L_{\text{reg},n}(j,z)=\frac{1}{N_n(A)}\sum_{i=1}^{n}(Y_i-\bar{Y}_A)^2\mathrm{I}(X_i\in A)-\frac{1}{N_n(A)}\sum_{i=1}^{n}\left((Y_i-\bar{Y}_{A_L})^2\mathrm{I}(X_i\in A_L)+(Y_i-\bar{Y}_{A_R})^2\mathrm{I}(X_i\in A_R)\right) \quad (2)$$

In this equation, the algorithm chooses the best split z^* along the feature X_j (e.g., task completion rate), where A is the current node, and A_L and A_R are the left and right child nodes, respectively. $\bar{Y}_A$, $\bar{Y}_{A_L}$, and $\bar{Y}_{A_R}$ are the mean performance scores in each region, and the algorithm selects the split that minimizes the error in performance prediction across the employee data. In terms of classification, the Gini impurity criterion is used to select the optimal split at each node. This ensures that the employee groups (e.g., high vs. low performers) become more homogeneous as we traverse deeper into the tree.

Predictive Performance and Generalization in Employee Evaluation

One of the key benefits of Random Forests in this employee performance evaluation is their ability to generalize well to new data. The generalization error of the Random Forest depends on two factors: the strength of individual trees and the correlation between trees. The predictive accuracy of the model is expressed in terms of the generalization error bound provided by Breiman (2001):

$$PE \leq \rho\frac{s}{(1-\rho)^2} \quad (3)$$

Here, s represents the strength of each individual decision tree (i.e., its predictive accuracy), and ρ is the correlation between the errors of different trees. In our employee performance evaluation, a higher s reflects a more accurate tree when predicting employee performance based on metrics like task completion and manager evaluations. A lower ρ indicates less correlation between trees, which leads to better ensemble predictions.

Application of Out-of-Bag (OOB) Error Estimation

In the employee performance dataset, Out-of-Bag (OOB) error estimation is used to assess the model's performance without requiring a separate validation set. Since each tree in the Random Forest is built using a bootstrap sample, approximately one-third of the data points are left out during the construction of any given tree. These OOB samples are used as test data to compute prediction errors. The OOB error is calculated by averaging the error across all trees, providing a robust estimate of the model's accuracy in predicting employee performance across different performance levels (high, medium, low). This OOB error serves as an unbiased estimate of the model's predictive performance, ensuring that the model is not overfitting to the training data, especially important when working with limited or imbalanced employee data.

Variable Importance in Employee Evaluation

Random Forests also allow us to measure the importance of various employee features in predicting performance outcomes. This is critical in determining which factors—such as task completion, attendance, peer feedback, or manager evaluations—have the greatest impact on predicting whether an employee is a high or low performer. The Mean Decrease in Impurity (MDI) for a feature X_j is computed by summing the reductions in the Gini impurity criterion (for classification tasks) or the mean squared error (for regression tasks) across all nodes in which the feature is used to split the data:

$$\mathrm{MDI}(X_j)=\frac{1}{M}\sum_{t=1}^{M}\sum_{s\in T_t,\text{split on }X_j}\frac{N_s}{N}\Delta i_s \quad (4)$$

In the employee performance context, the MDI measures how frequently a feature, such as task completion or peer feedback, contributes to reducing the model's prediction error. A higher MDI indicates a more influential feature, guiding HR professionals to focus on key performance indicators when assessing employees. The scalability of Random Forests is a major advantage in large-scale employee performance evaluation systems, especially when datasets involve numerous employees with varied performance metrics. The method remains consistent under certain assumptions, particularly in regression tasks, where the algorithm adapts well to high-dimensional data. As shown by Scornet et al. (2015), Random Forests are consistent estimators, meaning that as the size of the employee dataset increases, the model's predictions become more accurate.

Connecting the Insights to Employee Performance Evaluation

By grounding the application of the Random Forest algorithm in the context of employee performance, we can highlight how each mathematical component of the algorithm—such as bootstrap sampling, variable importance measures, and CART-split criteria—contributes to accurate predictions in the workplace setting. The use of Random Forests in this study allows us to leverage complex, high-dimensional employee data and extract meaningful insights into which performance metrics are most predictive of success, while ensuring generalization and robustness of the model's predictions. This integration of Random Forests with employee data not only provides a powerful predictive tool but also offers practical insights into employee performance drivers, making it valuable for HR decision-making and employee development strategies. By weaving these connections between the mathematical foundations of Random Forests and the specific application to employee performance evaluation, you create a narrative that demonstrates both the theoretical rigor of the algorithm and its practical utility in your research.

Feature Importance and Interpretability

One of the most powerful aspects of Random Forest is its ability to provide insights into feature importance. In performance evaluation, understanding which features (e.g., task completion rates, peer evaluations, or training participation) contribute the most to an employee's predicted performance can be highly valuable for decision-makers. The algorithm ranks the importance of each feature, giving a clear picture of which factors are driving the performance outcomes. In the context of employee performance, this interpretability is critical. For instance, if task completion rates are shown to be a significant predictor of overall performance, managers can focus on improving task efficiency across teams. Similarly, if peer feedback is a strong indicator of performance, organizations may consider enhancing collaborative efforts within teams. Random Forest thus not only provides predictive power but also actionable insights that can help improve organizational performance.

In practice, Random Forest can be applied to a wide array of employee performance data. Consider a dataset that includes task completion rates, manager feedback, peer evaluations, and attendance records. The Random Forest model can classify employees into different performance bands (e.g., high, medium, low performers) or predict future performance levels based on these inputs. For example, an HR department might use Random Forest to predict which employees are likely to excel or struggle in the coming year based on historical data. These predictions can guide decisions about promotions, training, or additional support. Random Forest's ability to handle both quantitative data (such as task completion rates) and qualitative data (such as feedback from managers and peers) makes it a versatile tool in employee performance evaluation.

Performance Evaluation Metrics for Random Forest

When evaluating the performance of the Random Forest model, several metrics can be considered depending on whether the task is classification or regression. For classification tasks, accuracy, precision, recall, and F1 score are commonly used to assess how well the model is predicting employee performance bands. For regression tasks, where the goal might be to predict an exact performance score, metrics like mean absolute error (MAE) or root mean square error (RMSE) can be applied. These metrics help to ensure that the model is not only making accurate predictions but also doing so in a way that aligns with the business's goals. For example, if the focus is on identifying high-performing employees for promotion, precision in predicting the top performers may be prioritized over overall accuracy.

Advantages of Using Random Forest in HR Analytics

Random Forest provides several key advantages when applied to employee performance evaluation. First, its ability to handle a large number of input variables without significant overfitting makes it ideal for performance prediction, where multiple factors may influence outcomes. Second, the algorithm's robustness in handling missing data ensures that HR departments can still make reliable predictions even when employee records are incomplete or contain missing values. Additionally, the scalability of Random Forest means that it can be applied to both small and large organizations. As companies grow and gather more performance-related data, the model can scale accordingly, providing consistent and accurate predictions. This adaptability is crucial for modern businesses, where data is continuously being generated and decision-makers require tools that can keep pace with this growth.

DATASET AND METHODOLOGY

Dataset Overview

In this study, we utilized a comprehensive dataset that includes key performance metrics to evaluate employee contributions across various departments. The dataset encompasses a range of indicators, including productivity rates, task completion rates, promotion readiness, work quality scores, and collaboration indices, offering a holistic perspective on employee performance. The Productivity Rate (%) and Task Completion Rate (%) measure how efficiently employees complete tasks and meet deadlines. Work Quality and Collaboration assess task precision and teamwork. Promotion Readiness (%) reflects potential for career growth, while Skill Development, Engagement, Time Management, and Leadership scores provide insights into personal growth and leadership abilities. Together, these metrics offer a comprehensive view of employee performance.

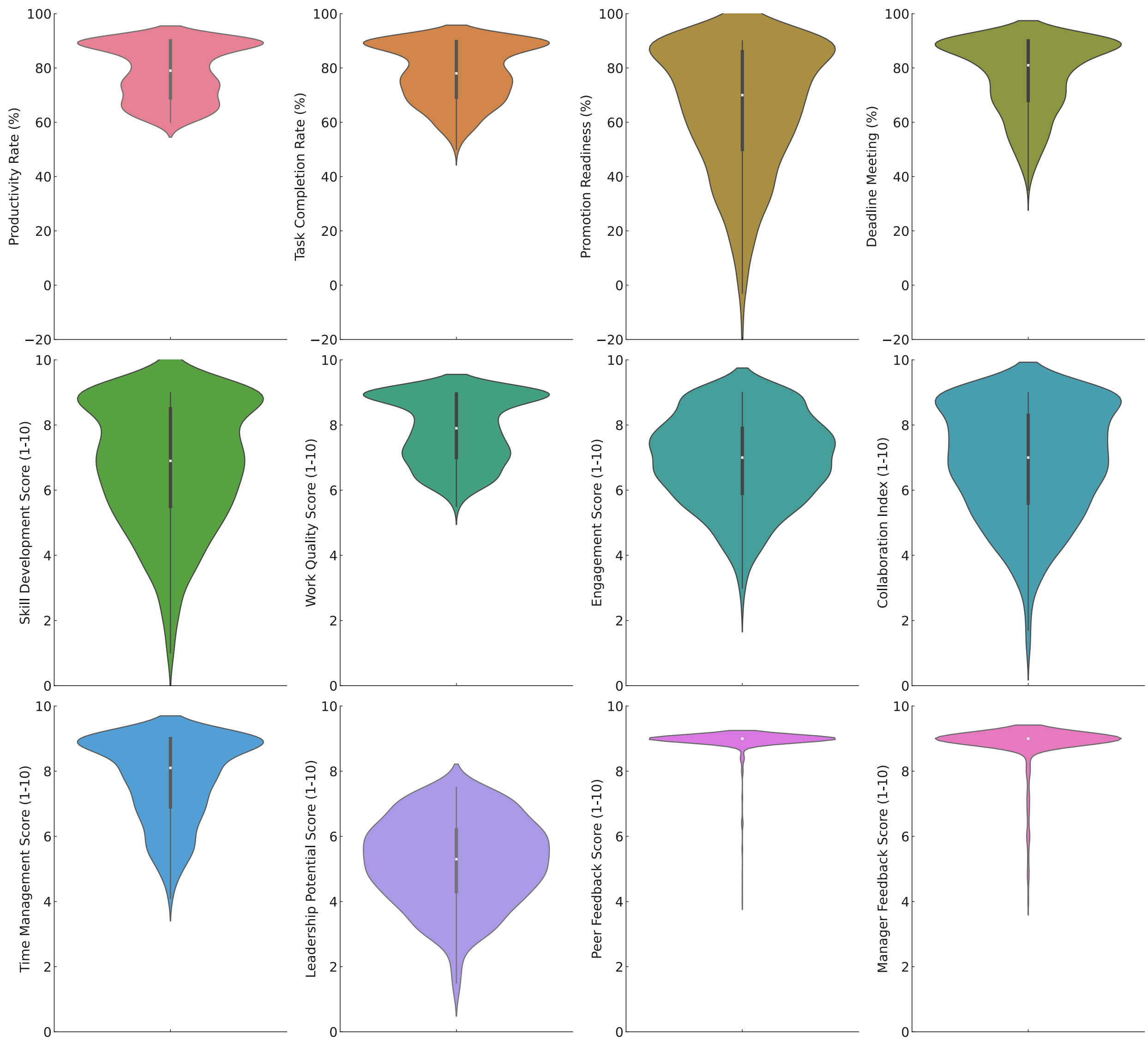

Figure 1 Distribution of Employee Performance Metrics

Figure 1 presents the distribution of these key performance metrics through violin plots, which visualize the spread and density of performance scores across employees in various departments. The top row of the figure illustrates percentage-based metrics, including Productivity Rate (%), Task Completion Rate (%), Promotion Readiness (%), and Deadline Meeting (%), with y-axis values ranging from -20 to 100 to capture a wide spectrum of performance levels. The second and third rows feature scaled scores (0 to 10) for qualitative metrics such as Work Quality, Collaboration, and Peer/Manager Feedback.

The violin plots in Figure 1 provide a visual representation of the variability and concentration of employee performance across departments. Wider sections in each plot indicate higher density of values, while narrower sections represent fewer occurrences. This allows for easy identification of performance trends, highlighting areas where employees excel or need improvement.

The dataset integrates feedback from multiple sources, ensuring a comprehensive perspective on employee performance. Structured reviews conducted by managers and team leads assess task-specific metrics like productivity, work quality, and adherence to deadlines. Customer feedback, especially for customer-facing roles, adds an external perspective, evaluating how well employees handle real-world interactions. Peer evaluations further complement this by offering insights into teamwork and collaboration, helping to measure employees' ability to positively contribute to group dynamics. By combining data from managers, peers, and customers, the dataset provides a balanced, multi-dimensional view of employee performance, allowing for targeted development and performance management strategies.

Data Processing and Analysis

Before applying machine learning algorithms to the dataset, several essential preprocessing steps were carried out. The first task was data cleaning, which involved dealing with missing values, duplicates, and outliers. Missing data was handled using imputation techniques or, where appropriate, certain entries were removed if their absence did not affect the overall integrity of the dataset. Duplicate entries were removed to ensure that no redundancy distorted the results. Outliers were carefully examined, and based on their influence, either removed or normalized to reduce their impact on model predictions. This cleaning process was critical for ensuring the dataset was accurate and reflective of real performance trends.

Another important preprocessing step was feature selection. Although the dataset included a wide range of metrics, only the most relevant features were retained for further analysis. For example, core performance indicators such as productivity, task completion rates, work quality, and collaboration scores were prioritized, while less significant features were excluded to reduce dimensionality and improve model performance. Additionally, normalization and scaling were applied to standardize the data, especially since different metrics like productivity (absolute numbers) and task completion rates (percentages) operated on different scales. Ensuring that all data points were on a common scale improved the performance of the machine learning models used in this study.

After preprocessing, machine learning algorithms were applied to the dataset to uncover patterns and make predictions about employee performance. One of the primary algorithms used was decision trees, a supervised learning method that classifies employees into different performance categories, such as high, medium, or low performers. Decision trees work by splitting the data based on key features that most effectively separate employees into these categories. This approach provides insight into which factors—whether productivity, work quality, or collaboration—are the strongest predictors of employee success. In addition to decision trees, clustering algorithms were employed to group employees based on similar performance characteristics. By identifying patterns among employees, clustering helps pinpoint groups that may need additional support or training, or who excel in certain areas like task completion but may struggle with collaboration.

To evaluate the effectiveness of the machine learning models, performance metrics such as accuracy, precision, and recall were used. Accuracy measures the proportion of correct predictions made by the model, giving a general sense of the model's effectiveness in classifying employee performance. However, accuracy alone may not provide the full picture, especially in datasets with imbalanced classes (e.g., few high performers). In these cases, precision and recall offer more detailed insights. Precision focuses on the proportion of true positive predictions (e.g., high performers correctly identified by the model), providing a clearer view of how reliable the model is in identifying specific performance categories. Recall, on the other hand, measures the proportion of actual positive cases that the model successfully identifies. This metric is especially important when the goal is to ensure that no high performers are overlooked by the model. To balance precision and recall, the F1 score is used as a combined measure, especially valuable when dealing with imbalanced data or when both precision and recall are of high importance.

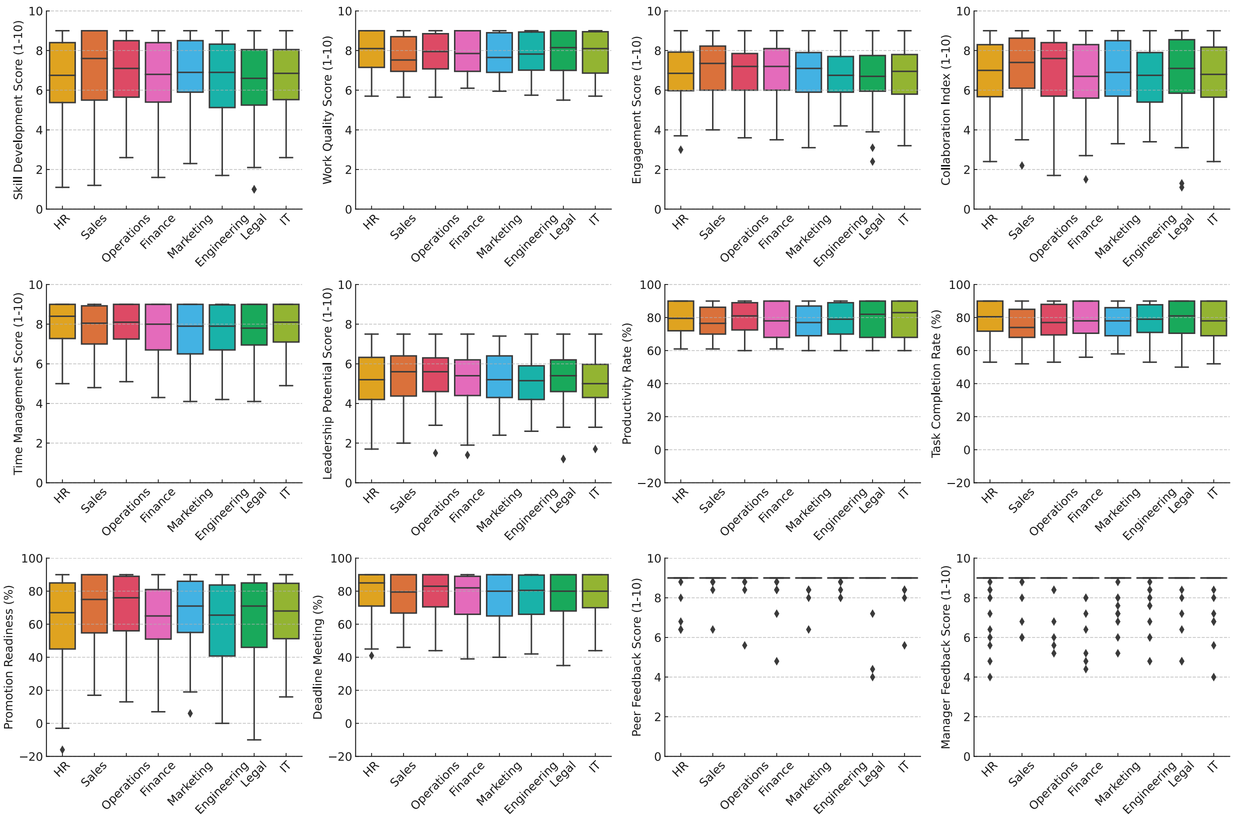

Figure 2 Distribution of Performance Metrics Across Departments as per ML feedback

EMPIRICAL RESULTS

Impact of ML Feedback on Employee Performance

The analysis of employee performance metrics across departments, as depicted in **Figure 2**, reveals significant variability in key areas such as productivity, work quality, and collaboration. By examining the distribution of these metrics, it becomes evident that certain departments consistently perform better in specific areas while others may have room for improvement. The box plots in Figure 2 show how departments vary in performance across metrics such as Productivity Rate (%), Task Completion Rate (%), Work Quality Score (1-10), and Collaboration Index (1-10). Departments like Marketing, Engineering, and Finance demonstrated higher productivity levels, as indicated by higher median scores in Productivity Rate (%) and Task Completion Rate (%). On the other hand, departments like HR and Sales showed more variability in these areas, suggesting a need for focused interventions to boost performance consistency.

Work quality also varied across departments, with departments like Operations and Finance performing well, as indicated by higher Work Quality Scores. This suggests that these departments have strong processes in place for maintaining the precision and thoroughness of their tasks. Collaboration scores reveal that certain departments, such as Engineering and Legal, could benefit from initiatives aimed at improving teamwork, as their Collaboration Index values are lower compared to departments like Marketing and Finance, which have higher median scores in this area.

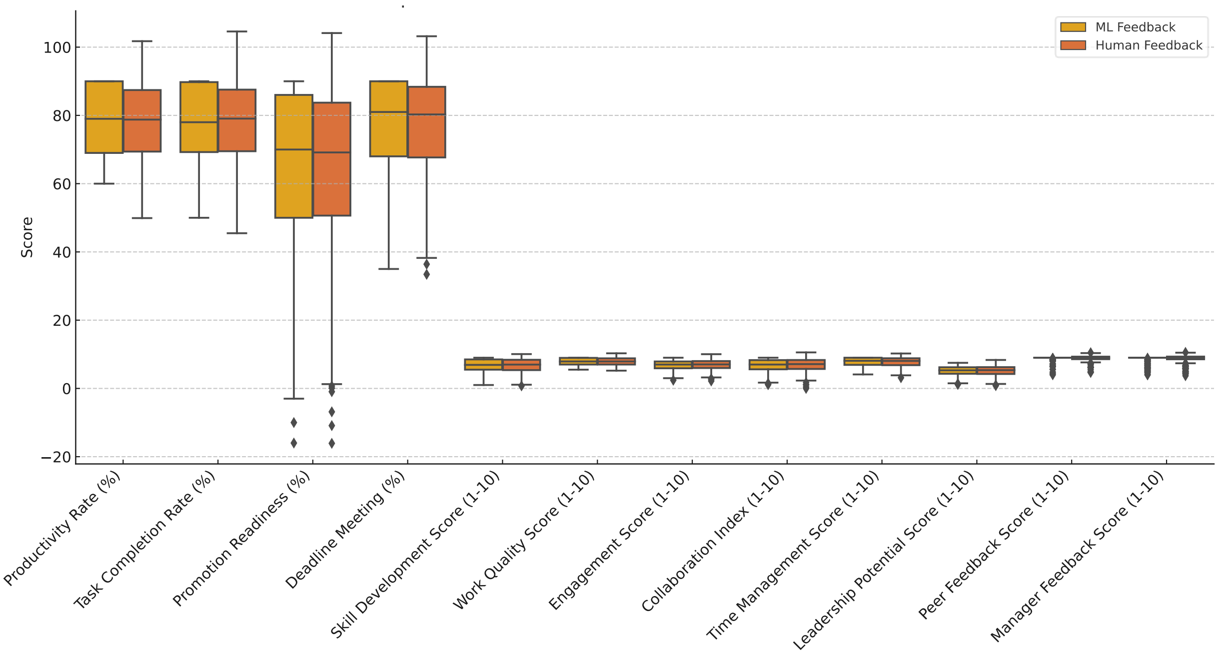

Figure 3 Comparison of ML and Human Feedback Across All Performance Metrics.

Comparison Between ML-Driven and Human Feedback

The comparison between ML-driven and human feedback, as demonstrated in **Figures 3** and **4**, highlights the distinct advantages and limitations of each approach in evaluating employee performance. ML-driven feedback, being highly objective and data-driven, provided more consistent and standardized evaluations across key metrics such as *Productivity Rate (%)*, *Task Completion Rate (%)*, and *Work Quality Score*. **Figure 4** illustrates this consistency, showing that ML feedback results in a more uniform distribution of performance scores, particularly for metrics that are easily quantifiable. In contrast, human feedback exhibited greater variability, which can often stem from subjective interpretations and personal biases.

In **Figure 3**, the scatter plot underscores the differences between ML and human feedback across all metrics, particularly in soft-skill areas such as *Leadership Potential* and *Collaboration Index*. While ML models rated employees based on clear data like task completion or productivity rates, human feedback considered more nuanced aspects, such as interpersonal skills, emotional intelligence, and decision-making abilities. These areas are harder for ML systems to quantify but are critical in assessing leadership potential and team dynamics. The scatter plot highlights where human evaluations diverged from ML, often capturing qualities that data-driven models may overlook.

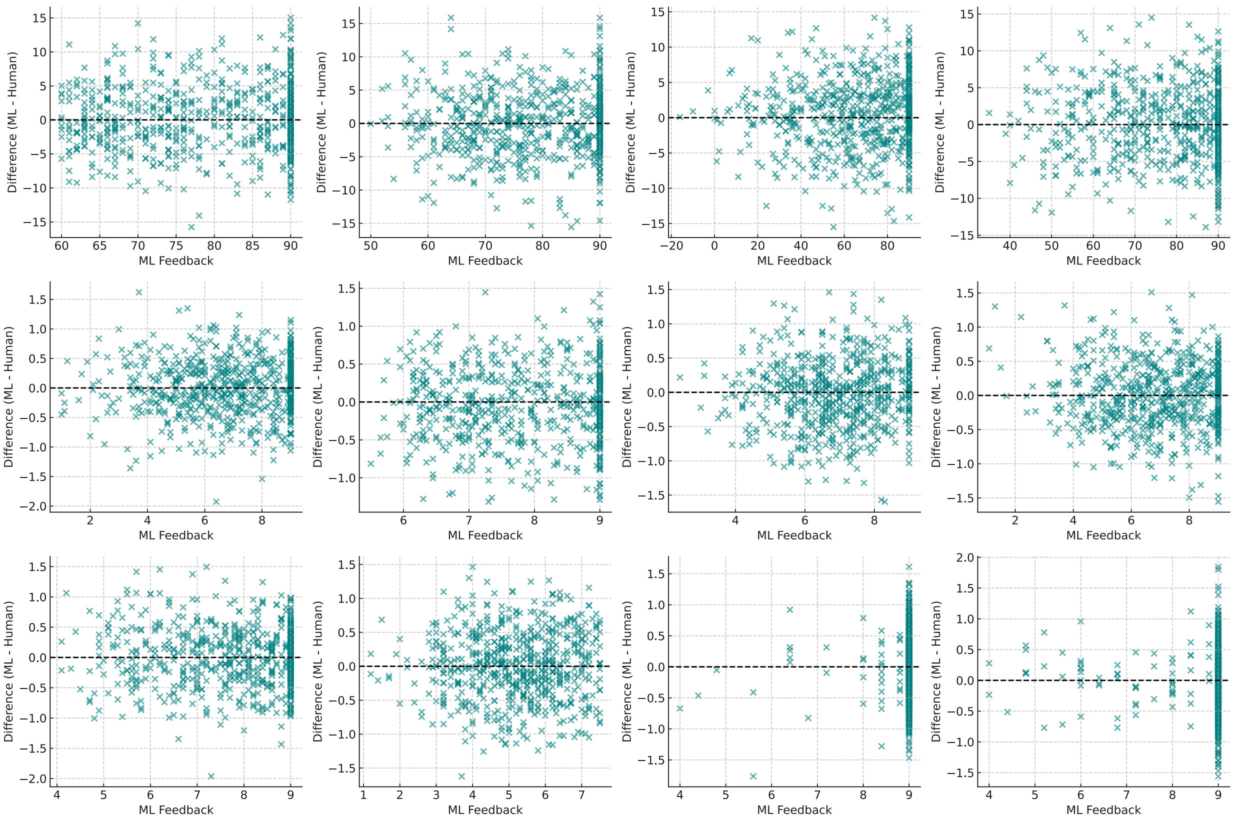

Figure 4 Difference Between ML and Human Feedback Across All Performance Metrics.

A key area where human feedback stood out was in leadership development. While ML models focused on measurable task-based performance, human evaluators offered deeper insights into traits like emotional intelligence, leadership charisma, and decision-making under pressure. Human feedback was more adept at recognizing leadership potential based on interactions and vision, qualities that are challenging for ML models to assess using structured data alone.

Conclusion

This study demonstrates the potential of integrating machine learning (ML) models, particularly random forests, with traditional human feedback systems to improve employee performance evaluations. ML-driven feedback provides more objective and consistent evaluations, especially for quantifiable metrics such as productivity, task completion, and work quality. Human feedback, on the other hand, remains crucial for evaluating qualitative aspects like leadership potential, emotional intelligence, and collaboration. The hybrid approach combining ML's consistency with human insights offers a more balanced and comprehensive evaluation system, addressing the limitations of both individual approaches.

While the results are promising, there are several limitations that must be addressed in future research. First, although ML systems reduce biases in quantifiable areas, they are limited in their ability to assess complex, interpersonal dynamics, such as leadership potential and emotional intelligence, which human evaluators often handle better. Future studies could explore more advanced ML models, including natural language processing (NLP) or deep learning, to better assess qualitative metrics. Additionally, the study's dataset is confined to specific performance metrics, and expanding the dataset to include a wider range of behavioral data, such as peer-to-peer interactions or customer feedback, could offer a more comprehensive view of employee performance.

Moreover, employee acceptance of ML-driven feedback varied across demographics. While technical and younger employees showed a higher acceptance of ML feedback, more experienced employees and those in leadership roles expressed skepticism, particularly when ML evaluations contradicted long-standing human judgments. Future work could explore ways to increase the adoption of ML feedback systems, perhaps by enhancing the explainability and transparency of ML models or by providing training on interpreting ML-driven insights alongside traditional feedback. Further research could also examine how to improve the integration of ML and human feedback in areas such as leadership development and succession planning, where qualitative judgments are more critical.

Reference

[1] P. Zbranek, “Data envelopment analysis as a tool for evaluation of employees performance,” Rev. Agric. Appl. Econ., vol. 16, no. 01, pp. 12–21, Feb. 2013.

[2] S. C. Payne, M. T. Horner, W. R. Boswell, A. N. Schroeder, and K. J. Stine-Cheyne, “Comparison of online and traditional performance appraisal systems,” J. Manag. Psychol., vol. 24, no. 6, pp. 526–544, Aug. 2009.

[3] M. H. Bazerman, R. I. Beekun, and F. D. Schoorman, “Performance evaluation in a dynamic context: A laboratory study of the impact of a prior commitment to the ratee,” J. Appl. Psychol., vol. 67, no. 6, pp. 873–876, Dec. 1982.

[4] N. Kaur and S. K. Sood, “A game theoretic approach for an IoT-based automated employee performance evaluation,” IEEE Syst. J., vol. 11, no. 3, pp. 1385–1394, Sep. 2017.

[5] G. Kabir, R. S. Sumi, R. Sadiq, and S. Tesfamariam, “Performance evaluation of employees using Bayesian belief network model,” Int. J. Manag. Sci. Eng. Manag., vol. 13, no. 2, pp. 91–99, Apr. 2018.

[6] S. Yang and Q. Zhu, “An evaluation model on employee performance based on improved BP neural network,” in 2008 4th International Conference on Wireless Communications, Networking and Mobile Computing, Dalian, China, 2008.

[7] P. Poyet and P. Haren, “Artificial intelligence modelling of complex systems,” in Modeling Techniques and Tools for Computer Performance Evaluation, Boston, MA: Springer US, 1989, pp. 265–290.

[8] A. Holzinger, “From machine learning to explainable AI,” in 2018 World Symposium on Digital Intelligence for Systems and Machines (DISA), Kosice, 2018.

[9] A. Holzinger, P. Kieseberg, E. Weippl, and A. M. Tjoa, “Current advances, trends and challenges of machine learning and knowledge extraction: From machine learning to explainable AI,” in Lecture Notes in Computer Science, Cham: Springer International Publishing, 2018, pp. 1–8.

[10] K. M. Kacmar, B. J. Collins, K. J. Harris, and T. A. Judge, “Core self-evaluations and job performance: the role of the perceived work environment,” J. Appl. Psychol., vol. 94, no. 6, pp. 1572–1580, Nov. 2009.

[11] K. Moh’dAlhendawi and A. Suhaimi Baharudin, “Proposing an optimization algorithm for employee competencies evaluation using artificial intelligence methods: Bayesian network and decision tree,” Int. J. Comput. Appl., vol. 71, no. 5, pp. 18–22, Jun. 2013.

[12] Y. Reich and S. V. Barai, “Evaluating machine learning models for engineering problems,” Artif. Intell. Eng., vol. 13, no. 3, pp. 257–272, Jul. 1999.

[13] T. R. Manoharan, C. Muralidharan, and S. G. Deshmukh, “An integrated fuzzy multi-attribute decision-making model for employees’ performance appraisal,” Int. J. Hum. Resour. Manag., vol. 22, no. 3, pp. 722–745, Jan. 2011.

[14] S. Halperin and U. Zwick, “Optimal randomized EREW PRAM algorithms for finding spanning forests,” J. Algorithm., vol. 39, no. 1, pp. 1–46, Apr. 2001.

[15] E. M. Rodrigues, M.-F. Sagot, and Y. Wakabayashi, “Some approximation results for the maximum agreement forest problem,” in Approximation, Randomization, and Combinatorial Optimization: Algorithms and Techniques, Berlin, Heidelberg: Springer Berlin Heidelberg, 2001, pp. 159–169.

[16] P. Latinne, O. Debeir, and C. Decaestecker, “Limiting the number of trees in random forests,” in Multiple Classifier Systems, Berlin, Heidelberg: Springer Berlin Heidelberg, 2001, pp. 178–187.

[17] M. H. Shaker and E. Hüllermeier, “Aleatoric and epistemic uncertainty with random forests,” arXiv [cs.LG], 03-Jan-2020.