



Computational Modeling Using Linear Regression and Random Forest to Analyze the Impact of Workload on Employee Performance Evaluations

Safitri Jaya^{1*}, Aries Yulianto², Edi Purwanto³

¹ Department of Informatics, Universitas Pembangunan Jaya, South Tangerang 15413, Indonesia

² Department of Psychology, Universitas Pembangunan Jaya, South Tangerang 15413, Indonesia

³ Department of Management, Universitas Pembangunan Jaya, South Tangerang 15413, Indonesia

Corresponding Author Email: safitri.jaya@upj.ac.id

Copyright: ©2025 The authors. This article is published by IIETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/ijcmem.130202>

ABSTRACT

Received: 30 December 2024

Revised: 23 March 2025

Accepted: 30 March 2025

Available online: 30 June 2025

Keywords:

performance evaluation, workload, linear regression, Random Forest

Each employee has a number of workloads in the form of tasks and responsibilities that must be completed within a certain period of time. There are several aspects that are assessed in performance evaluations, such as the number of working hours per week, the number of projects handled, the number of overtime hours to complete work, the number of sick days, the number of team members, the number of hours to develop self-skills, job promotion offers, etc. All of these aspects certainly impact performance scores, employee satisfaction scores, and the ability to survive. Excessive workload will have a negative impact on physical and mental health, performance, and employee satisfaction levels. This study aims to analyze the results of employee performance evaluations based on workload factors using machine learning approaches such as linear regression and random forest. The computational results will be used to compare the effectiveness of machine learning models and analyze the accuracy of the assessment results. The significance of this study lies in its potential to enhance employee performance management systems by providing accurate, data-driven insights for decision-making processes such as promotions, compensation, and workforce planning. Practical and fair employee performance assessments will enable decision-makers to make informed choices regarding job promotions, salary increases, annual bonuses, and employee career development.

1. INTRODUCTION

The sustainability of an organization is closely linked to the performance of its employees [1]. Current employee performance evaluation systems require re-evaluation due to the influence of external factors such as social, economic, and environmental conditions that can distort the assessment process [2, 3]. A fair and accurate evaluation system can assist decision-makers in making ethical and practical choices regarding employee promotions, career advancement, and training needs [4]. To achieve long-term cooperation and growth, companies must guide employees toward understanding the organizational vision and develop their strategic management capabilities [5]. In this context, career planning becomes essential as job competition intensifies [5]. Organizations prioritizing identifying and nurturing employee talents are better equipped for sustainable development [6]. This study focuses on employees in the banking sector in Indonesia, where workload indicators such as the number of transactions processed, customer accounts managed, compliance reporting, and hours of training significantly influence employee performance evaluations. Employee performance identification often hinges on organizational structure, strategic work unit dynamics, and the core

competencies required to complete assigned tasks and responsibilities [7].

Employee job satisfaction and work behavior are pivotal in determining organizational success, especially in industrial settings [8]. Factors such as employee commitment, workplace stress [9], absenteeism, and turnover [10] are significantly influenced by workload satisfaction [11]. Workload management affects productivity, employee engagement in problem-solving, and employees' willingness to exert effort to improve performance [12]. The alignment between employee competence and job requirements is key to performance and satisfaction [13]. For example, assigning a highly analytical employee to repetitive administrative tasks may lead to disengagement, while placing a technically underprepared employee in a high-responsibility role can result in errors, stress, and dissatisfaction. Such mismatches can compromise both individual outcomes and organizational goals. Previous studies highlight the importance of achieving this alignment through strategic interventions that leverage the mediating role of employee performance and satisfaction and the moderating influence of social exchange within the workplace [14-17]. Such strategic interventions may include targeted upskilling programs for employees with competency gaps, job rotation schemes to align skills with tasks better, and

personalized coaching or mentorship initiatives to enhance role clarity and support performance development, particularly in high-pressure sectors such as banking.

Despite substantial research, gaps still need to be addressed in the comprehensive analysis of workload factors as predictive tools for evaluating employee performance. Existing studies underscore the necessity of considering employee competence and job alignment in developing performance improvement strategies. However, integrating these insights with advanced methodologies such as machine learning still needs to be improved. Machine learning offers the potential to uncover complex patterns and relationships within workload factors, enabling more informed, data-driven decision-making. In particular, this study employs linear regression to identify and quantify the strength of individual predictors and Random Forest algorithms to capture non-linear interactions, thereby assessing the relative importance of multiple workload variables simultaneously. These techniques enable a more comprehensive examination of the multidimensional nature of employee performance, which traditional statistical methods may not fully capture. This study aims to address these gaps by analyzing employee performance evaluation results through the lens of workload factors, utilizing machine learning approaches to identify key patterns and provide actionable insights for enhancing individual and organizational performance. The remainder of this paper is organized as follows: Section 2 is a literature review. Section 3 describes the methodological framework, including the application of linear regression and Random Forest models. Section 4 presents the computational results and analysis. Section 5 offers a detailed discussion of the findings and their implications, and Section 6 concludes with final insights and recommendations for future research.

2. LITERATURE REVIEW

2.1 Employee performance evaluation

Employee performance evaluation remains a central concern in human resource management due to its direct impact on organizational outcomes. Traditional appraisal methods typically rely on supervisor ratings, behavioral assessments, and predefined output metrics [1, 2]. However, such systems often fail to account for contextual variables like workload intensity or employee competence alignment, potentially leading to biased or incomplete evaluations [3, 4].

2.2 Impact of workload on performance and well-being

A growing body of research highlights the impact of workload on employee outcomes. Excessive workload has been linked to reduced job satisfaction, physical and mental fatigue, and increased intention to turnover. For instance, Holden et al. [9] demonstrated that high nursing workloads compromise patient safety and employee well-being. In a study focused on medical professionals, Lu et al. [11] identified job stress and workload as predictors of physician attrition. In the banking sector, Lee and Migliaccio [18] noted that physically repetitive and high-pressure tasks decrease marginal productivity, especially when unaccompanied by role-specific training and rest periods.

2.3 Role of competency alignment and training

The alignment between employee competencies and job demands has a significant impact on performance and satisfaction. Riyanto et al. [13] found that unmet training needs and skill-job mismatches lead to reduced motivation and suboptimal outputs. Thi Nong et al. [14] confirmed that competence-job fit, moderated by social exchange relationships, enhances both business performance and employee engagement. Structured training programs, mentorship, and career development initiatives help mitigate these risks by enabling continual learning and adaptability, particularly vital in dynamic sectors such as banking [19].

2.4 Machine learning applications in human resource analytics

As organizations increasingly digitize HR practices, machine learning has emerged as a valuable tool for analyzing workforce dynamics. Linear regression models are well-suited for estimating the influence of individual workload components on performance outcomes, offering interpretability and hypothesis testing [20]. In contrast, Random Forest models provide higher predictive accuracy by identifying complex, nonlinear patterns across multidimensional datasets [21, 22]. These approaches have been successfully applied to areas such as turnover prediction, recruitment filtering, and unbiased performance scoring [4, 23].

2.5 Research gap and study contribution

Although prior research addresses workload-performance relationships and introduces machine learning in HR contexts, systematic integration of these models with real workload datasets, especially in the banking industry, remains limited. This study contributes by filling that gap through a comparative computational analysis of employee workload data using both linear regression and Random Forest techniques. It aims to enhance the fairness, transparency, and predictive power of performance evaluation frameworks.

3. METHOD

3.1 Linear regression in machine learning

Machine Learning, a subset of Artificial Intelligence, specializes in creating algorithms and statistical models designed to learn from data and make predictions [24]. Among these algorithms is linear regression, a supervised machine-learning technique. It analyzes labeled datasets, identifies patterns, and fits data points to the most optimized linear functions [23]. These functions can then be applied to make accurate predictions on new datasets.

It is essential to understand supervised machine learning algorithms. This type of machine learning involves training an algorithm using labeled data, where the target values of the dataset are already known [25, 26]. Supervised learning is further divided into two categories: (1) Classification. This approach predicts the class or category of a dataset based on independent input variables [27]. Classes represent categorical or discrete values. For example, determining whether an image depicts a cat or a dog. (2) Regression. This method predicts

continuous output values based on independent input variables [27, 28]. For instance, forecasting house prices using factors such as the property's age, proximity to the main road, location, and size.

Linear regression is a supervised machine learning algorithm that models the linear relationship between a dependent variable and one or more independent features by fitting a linear equation to the observed data [20]. When there is a single independent feature, it is referred to as Simple Linear Regression, while the presence of multiple features is known as Multiple Linear Regression. Beyond its predictive tool role, linear regression is a foundational framework for numerous advanced models. Methods such as regularization and support vector machines build upon its principles, broadening its applications [29]. Furthermore, linear regression is critical in assumption testing, allowing researchers to evaluate and validate fundamental assumptions about their data.

3.2 Simple linear regression

This is the most basic type of linear regression, involving a single independent variable and a single dependent variable. The equation for simple linear regression is:

$$y = a + bx \quad (1)$$

where:

- y is the dependent variable
- x is the independent variable
- a is the intercept
- b is the slope

The main goal of linear regression is to identify the best-fit line, which minimizes the error between the predicted and actual values. The best-fit line is characterized by having the most possible minor mistake. As its equation describes, this line represents the relationship between the dependent and independent variables. The slope of the line reflects the extent to which the dependent variable changes for each unit change in the independent variable. Mathematically, the best-fit line is determined using the Ordinary Least Squares (OLS) method, which minimizes the sum of squared residuals, the differences

between the observed values and the values predicted by the linear model. The OLS method calculates the intercept (a) and slope (b) coefficients by solving for the values that reduce the overall prediction error across all data points. This approach ensures that the resulting line is optimally positioned to represent the general trend in the data. This relationship is illustrated in Figure 1, which shows how changes in the independent variable (X) correspond to predicted changes in the dependent variable (Y).

In this context, Y is referred to as the dependent or target variable, while X is the independent variable, also known as the predictor of Y. Various functions or models can be used for regression, with a linear function being the simplest. Here, X can represent either a single feature or multiple features related to the problem. Linear regression predicts the value of the dependent variable (Y) based on the given independent variable (X), which is why it is called Linear Regression. In the example, X (input) represents work experience, and Y (output) corresponds to a person's salary. The regression line shown is the best-fit line for the model.

3.3 Random Forests in machine learning

The Random Forest algorithm is a robust and versatile tree-based Machine Learning technique [21]. It operates by constructing multiple Decision Trees during the training phase. Each tree is built using a random subset of the dataset, and a random subset of features is evaluated at each split [30]. This element of randomness introduces diversity among the trees, which helps to mitigate overfitting and enhances the model's overall predictive performance [22].

During the prediction phase, the algorithm combines the outputs of all the trees. It uses majority voting for classification tasks while calculating the average of the trees' predictions for regression tasks. This ensemble approach leverages the collective insights of the trees, resulting in stable and accurate outcomes. Random Forests are widely applied to both classification and regression problems and are celebrated for their ability to manage complex datasets, minimize overfitting, and deliver reliable predictions across various domains. This concept is illustrated in Figure 2.

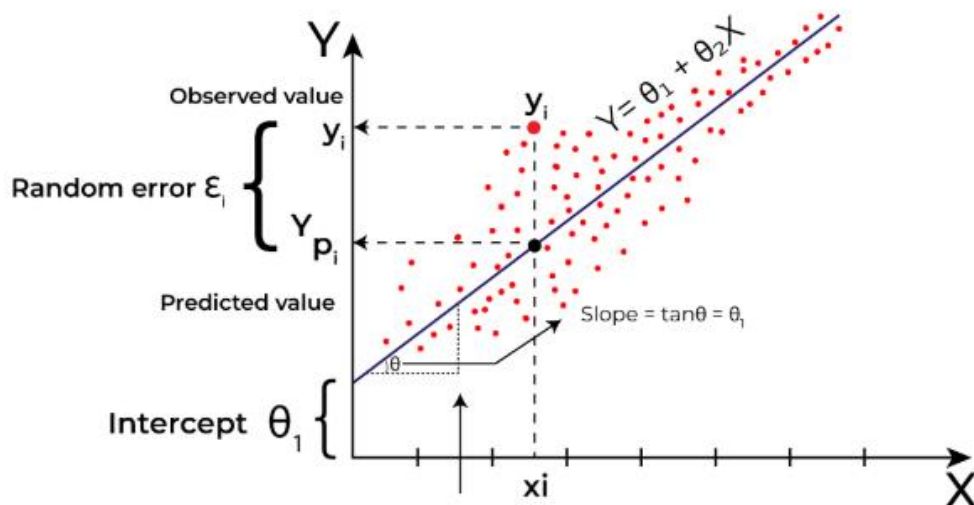


Figure 1. An illustration of linear regression

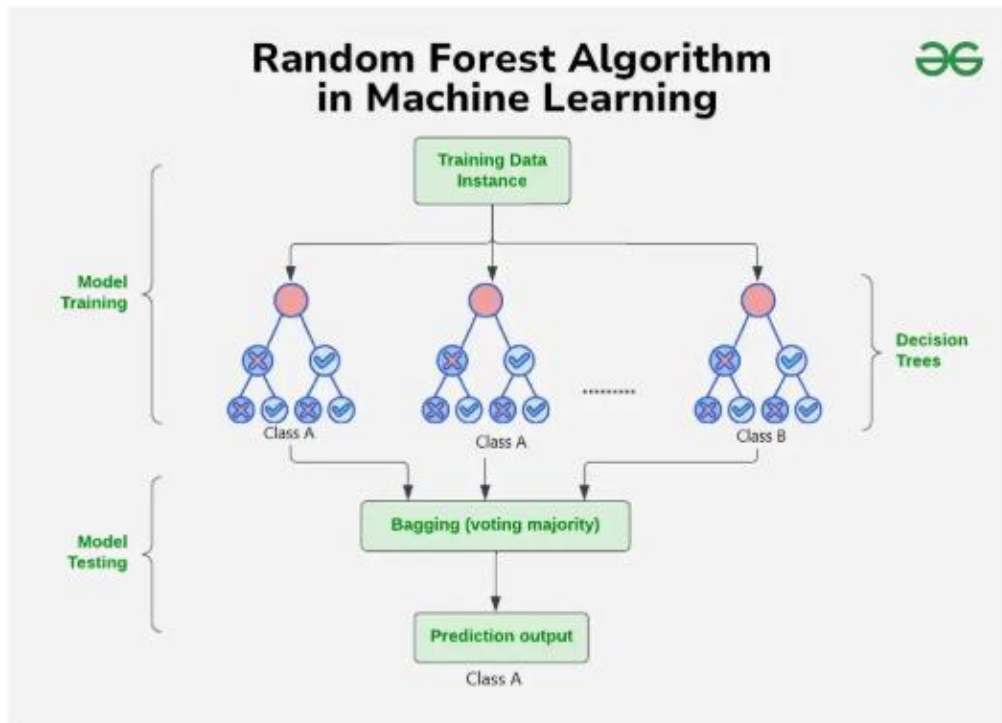


Figure 2. An illustration of a Random Forest

3.4 Computational modeling

This research's linear regression computational modeling was developed using the Python programming language. The modeling process stages consist of several steps, as shown in Figure 3, with the following details:

1. **Data Preparation:** Data consists of several variables that will be processed. Determine which is the dependent variable and which is the independent. Data is saved in file form using .csv format. The dataset used in this study was collected from an internal human resource database of a mid-sized commercial bank in Indonesia, covering employee performance records from 2022 to 2023. It includes 100,000 anonymized employee profiles and variables such as working hours per week, number of projects handled, overtime hours, sick days, number of team members, hours spent on training, promotion offers, employee satisfaction scores, and retention status. The performance score serves as the dependent variable. Data confidentiality and ethical approval were maintained throughout.
2. **Data Division:** Data can be divided into two parts, namely training data and test data. Training data is used to train the model, while test data is used to test the model's performance.
3. **Model Building:** The next step is to create a model using a linear regression algorithm after the data is divided. This model will learn from training data to understand the relationship between features and home prices.
4. **Model Training:** Models are trained using training data to learn patterns in the data and to help understand structures or relationships in complex data, which may be difficult to understand manually. This can help in better decision making or the development of new insights. The goal is to produce models that can perform these tasks with good accuracy and performance on new, never-before-seen data.

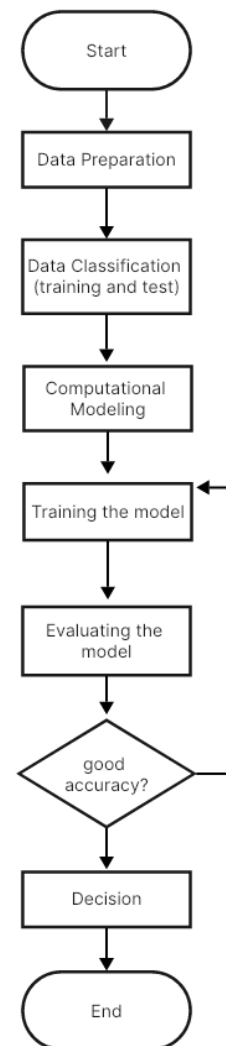


Figure 3. An illustration of computational modeling

5. Model Testing: Once the model is trained, we test its performance using test data that the model has never seen before. This aims to evaluate how well the model can predict the price of a house based on its features.
6. Model Evaluation: Finally, we evaluate the model performance by measuring the prediction accuracy. This accuracy shows how well the model can predict the price of a home based on its features. The higher the accuracy value, the better the model performance. If the model created can still not produce good predictions, we need to readjust the model.

4. RESULT

Here are some of the results obtained through machine learning computing and other statistical analysis.

4.1 Correlation matrix

Figures 4 and 5 show the correlation matrix of all features. Performance score is the dependent variable, while hours worked per week, projects handled, overtime hours, sick days, number of teams, training hours, promotions, employee satisfaction score, and retention are independent variables. All features show significant relationships in the analysis of employee performance appraisal.

4.2 Machine learning model training

Figure 6 shows the asymmetric size of the data distribution of values from all features. Based on the data modelling results, it is concluded that almost all features used in the analysis process show normal asymmetric sizes, except for the ability to survive the feature of 2.66. This number has a positive meaning, as the number of employees who have the ability to survive is greater than the number of employees who resign.

Of all the features, the highest correlation was between the dependent variable of work assessment and the independent variables of the number of projects handled, training hours, and number of teams. This provides the conclusion that the results of the performance assessment are influenced by these three factors. Figures 6, 7, and 8 show the distribution of the relationship values between work assessment and the independent variables in question.

The machine learning model training results with linear regression obtained an accuracy value of 81.32% and a precision of 83%. The results of Random Forest obtained an accuracy value of 93% and a precision of 90%.

Figure 9 uses machine learning modelling to show the performance measurement results for the employee performance evaluation classification problem. Performance evaluation is factored by the project's features, training hours, and the number of teams.

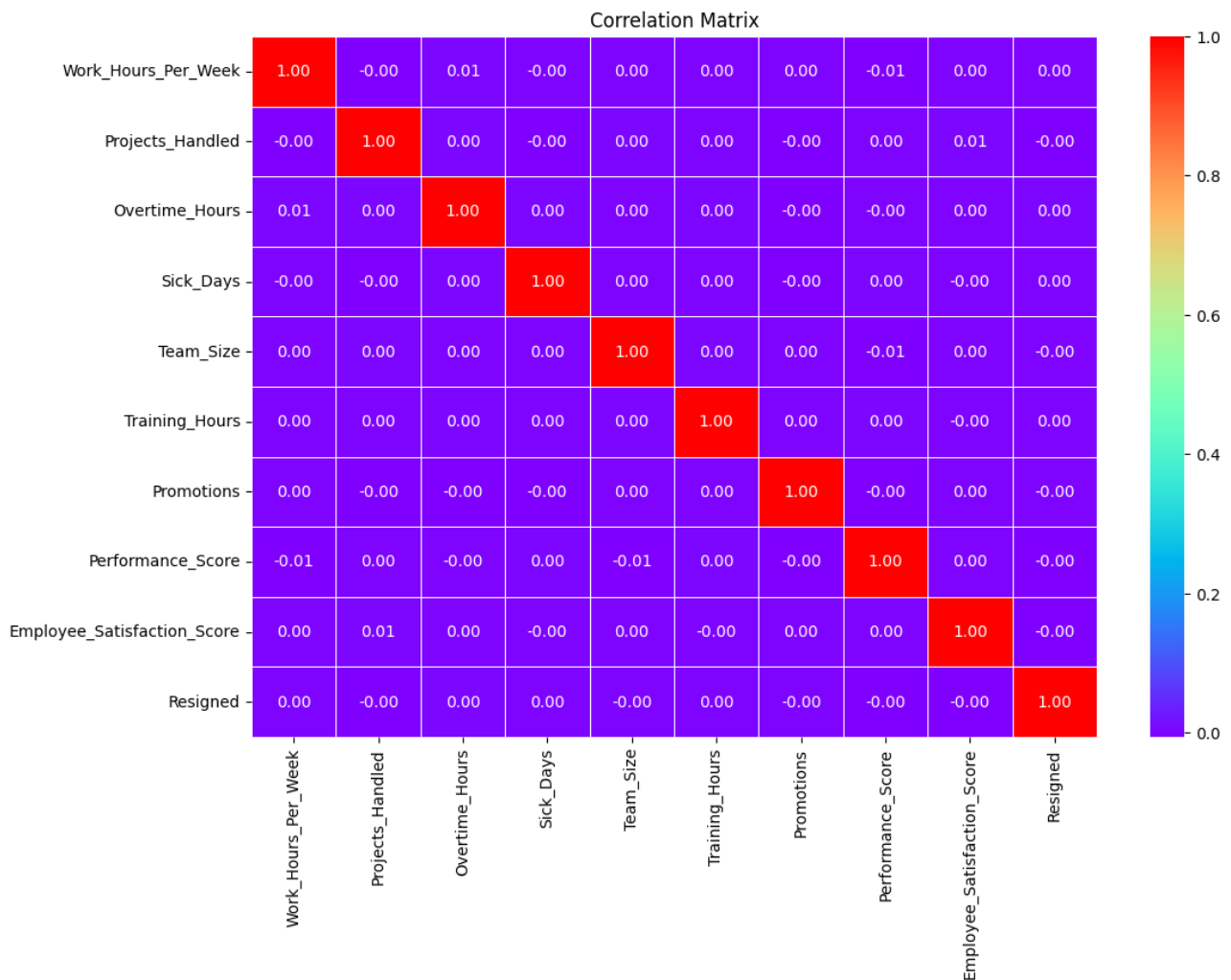
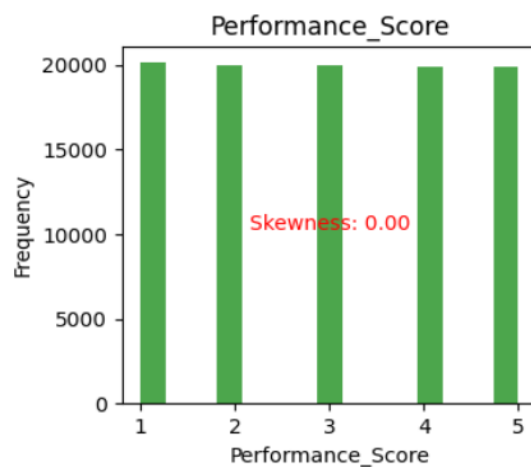
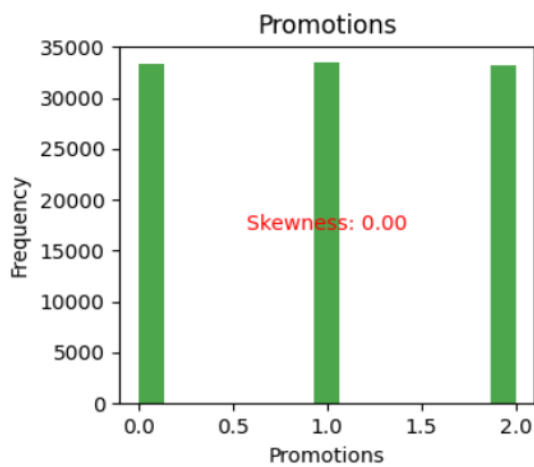
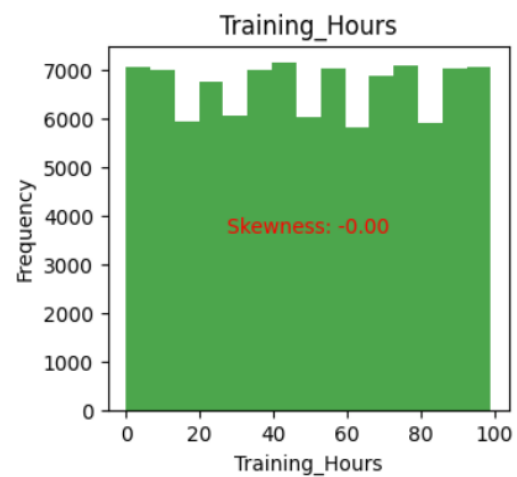
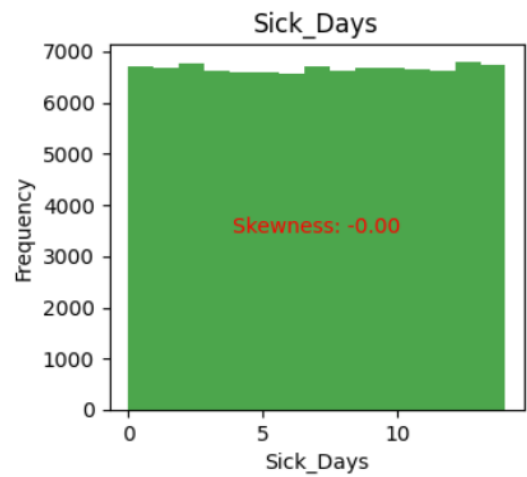
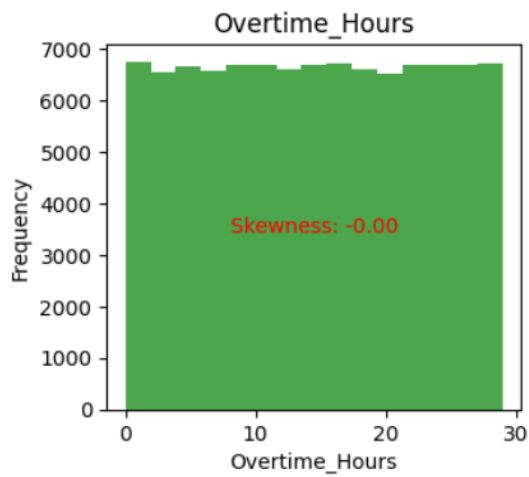
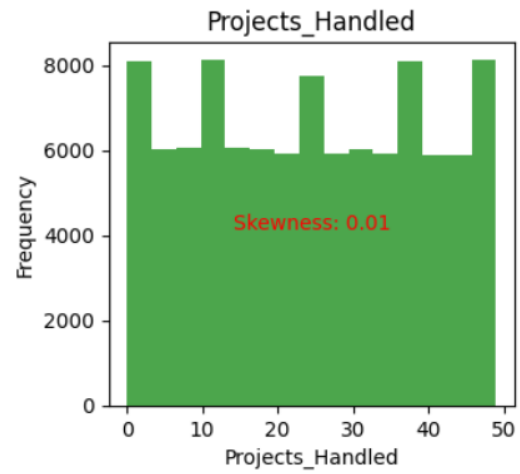
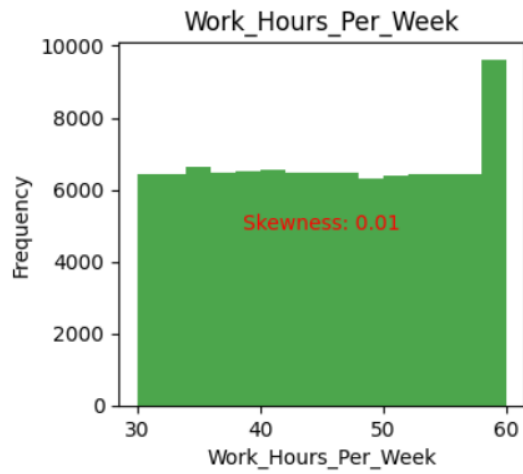


Figure 4. Correlation matrix



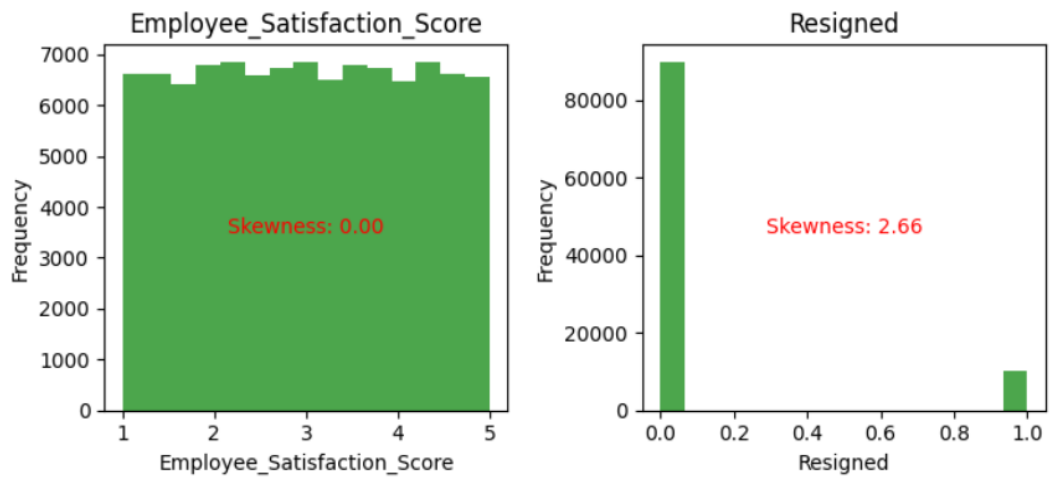


Figure 5. Skewness



Figure 6. Distribution of work value by projects handled



Figure 7. Distribution of work value by training hours

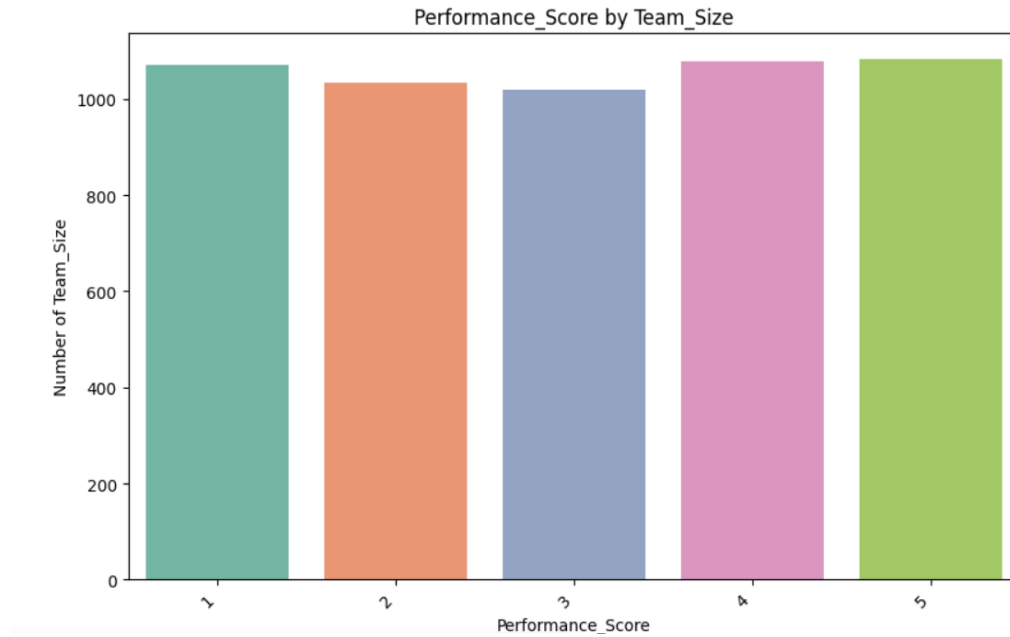


Figure 8. Distribution of work values by number of teams

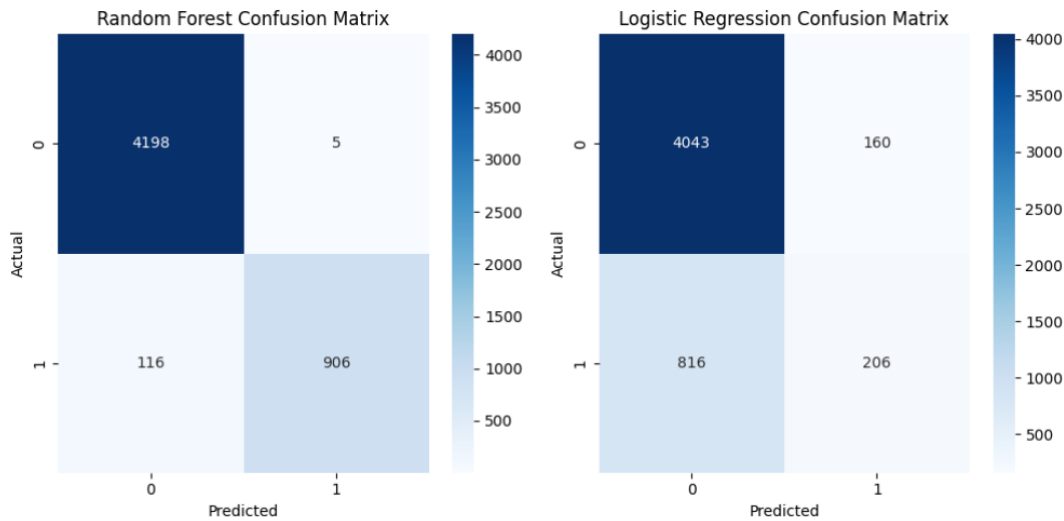


Figure 9. Performance evaluation results of the model prediction results

5. DISCUSSION

The findings from this study underscore the critical relationship between workload factors and employee performance evaluations. Using linear regression and Random Forest algorithms, this research demonstrates the potential of machine learning in predicting and analyzing key variables that impact performance. The results provide valuable insights into how specific workload aspects, such as the number of projects handled, training hours, and team size, significantly influence performance scores.

The correlation matrix revealed a strong relationship between these factors and employee performance, validating their importance in assessment frameworks. Among the features analyzed, the number of projects handled exhibited the highest correlation with performance outcomes, highlighting the importance of balancing workload distribution to optimize employee productivity. Training hours

also showed a substantial impact, emphasizing the role of continuous skill development in improving individual and organizational performance. Similarly, team size was identified as a significant variable, suggesting collaboration and team dynamics are crucial in achieving optimal results.

These findings complement existing literature, which explores various dimensions of employee performance and productivity, further enriching the understanding of these relationships. Workload distribution has been identified as a key determinant of employee outcomes, with Holden et al. [9] highlighting its importance in fostering job satisfaction and reducing burnout. Similarly, Lee and Migliaccio [18] emphasize the critical role of effective workload management in maintaining productivity levels, particularly in the construction industry. These studies collectively reinforce the view that balanced workload allocation is essential for achieving sustainable employee performance. Continuous skill development also emerges as a vital factor in enhancing

employee performance. Ismail et al. [31] found strong correlations between job satisfaction, organizational commitment, and performance outcomes, illustrating the benefits of investing in employee training. The importance of skill development becomes even more pronounced in dynamic work environments, as noted by Rojanasart et al. [19], who highlight the necessity of adapting to rapidly changing conditions. These findings underscore the value of ongoing training in building workforce resilience and effectiveness.

Team size and collaboration further contribute significantly to employee performance, as demonstrated by Holden et al. [9], who linked team dynamics to both employee and patient outcomes in healthcare settings. This finding underscores the broader implications of teamwork and collaboration across various industries. While the current study focuses on organizational factors such as workload, training, and team dynamics, other research highlights additional dimensions. For example, Boström et al. [32] examine individual health factors, such as musculoskeletal symptoms, which can substantially impact productivity. These insights suggest that enhancing employee performance requires a multifaceted approach, incorporating both organizational strategies and individual well-being. The interplay of workload management, skill development, and team dynamics serves as a cornerstone for improving employee performance and productivity. At the same time, recognizing the influence of individual health and well-being highlights the need for a holistic strategy that integrates organizational and personal considerations to achieve optimal outcomes.

The computational models utilized in this study further illustrate the advantages of integrating machine learning approaches into performance evaluations. Linear regression provided a transparent, interpretable framework for understanding how independent variables impact the dependent variable, while the Random Forest algorithm demonstrated superior accuracy and precision, achieving values of 93% and 90%, respectively. This highlights the robustness of ensemble methods in handling complex datasets and reducing overfitting risks compared to simpler models.

A notable strength of the Random Forest approach is its ability to capture non-linear relationships and interactions among variables, which may be overlooked by linear regression. However, linear regression remains invaluable for its simplicity and ease of providing actionable insights through model coefficients. These methods complement each other, offering a comprehensive toolkit for decision-makers aiming to refine performance evaluation systems.

The study's findings have practical implications for organizational management. By leveraging machine learning models, companies can enhance the fairness and accuracy of their performance evaluations, facilitating data-driven decisions regarding promotions, salary adjustments, and career development initiatives. Moreover, understanding the interplay between workload factors and performance can inform strategies to improve employee well-being and reduce turnover rates by aligning job demands with individual capabilities.

Future research should explore the integration of additional machine learning models to uncover deeper insights and validate the generalizability of these findings across diverse organizational settings. Expanding the dataset to include job satisfaction, workplace culture, and external socio-economic factors may provide a more holistic understanding of employee performance dynamics. Additionally, longitudinal

studies could help assess the long-term impact of workload adjustments on performance outcomes.

6. CONCLUSIONS

This study highlights the effectiveness of computational modeling in evaluating employee performance based on workload factors using machine learning approaches such as linear regression and Random Forest. The results demonstrate that key workload features, such as the number of projects handled, training hours, and team size, significantly influence performance outcomes. By leveraging these models, organizations can gain deeper insights into the relationships between workload and performance, enabling more informed and equitable decision-making.

Linear regression provides a straightforward and interpretable framework, offering apparent coefficients to quantify the impact of each independent variable. On the other hand, with its higher accuracy and precision, Random Forest excels at capturing complex, non-linear relationships, making it a valuable tool for robust performance evaluations. Together, these methods underscore the importance of integrating traditional statistical models with advanced machine learning techniques to enhance the predictive power of performance assessments.

The practical implications of this research are far-reaching. Organizations can apply these findings to develop fair and efficient workload management strategies, optimize employee satisfaction, and improve retention rates. Moreover, machine learning tools can support strategic decisions in promotions, salary adjustments, and career development, ensuring alignment between organizational goals and individual potential.

Future research should expand on this work by incorporating variables such as workplace culture, employee engagement, and external socio-economic factors. Longitudinal studies are also recommended to assess the sustained impact of workload adjustments over time, further refining the predictive capabilities of these computational models.

ACKNOWLEDGEMENT

This research was funded by a grant from the Center for Research and Community Service (LP2M) of Universitas Pembangunan Jaya, Indonesia. The support provided under contract number 012/PKS-P2M/UPJ/09.23 has been instrumental in completing this study. We extend our gratitude to LP2M for enabling this research.

REFERENCES

- [1] Purwanto, E. (2018). Moderation effects of power distance on the relationship between job characteristics, leadership empowerment, employee participation and job satisfaction: A conceptual framework. *Academy of Strategic Management Journal*, 17(1): 1-9.
- [2] Carmen, S., Assistant, P.H.D., Victoria, B. (2014). Study concerning the assessment and evaluation of human factor in SC hidroelectrica SA, curtea de arges. *Annals - Economy Series*, 4: 199-206.

- [3] Van Lerberghe, W., Adams, O., Ferrinho, P. (2002). Human resources impact assessment. *Bulletin of the World Health Organization*, 80(7): 525.
- [4] 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>
- [5] Montani, F., Odoardi, C., Battistelli, A. (2015). Envisioning, planning and innovating: A closer investigation of proactive goal generation, innovative work behaviour and boundary conditions. *Journal of Business and Psychology*, 30(3): 415-433. <https://doi.org/10.1007/s10869-014-9371-8>
- [6] Stopochkin, A., Sytnik, I., Wielki, J., Karaś, E. (2022). Transformation of the concept of talent management in the era of the fourth industrial revolution as the basis for sustainable development. *Sustainability (Switzerland)*, 14(14): 8727. <https://doi.org/10.3390/su14148727>
- [7] Vural, Y., Vardarlier, P., Aykir, A. (2012). The effects of using talent management with performance evaluation system over employee commitment. *Procedia - Social and Behavioral Sciences*, 58: 340-349. <https://doi.org/10.1016/j.sbspro.2012.09.1009>
- [8] Pourshahabi, V., Shahraki, H. (2024). Investigating the relationship between a safe and healthy work environment and counterproductive behavior of employees in the industry and mine bank. *Power System Technology*, 48(1): 2321-2332. <https://doi.org/10.52783/pst.501>
- [9] Holden, R.J., Scanlon, M.C., Patel, N.R., Kaushal, R., Kamisha, Escoto, H., Brown, R.L., Alper, S.J., Arnold, J.M., Shalaby, T.M., Murkowski, K., Karsh, B.T. (2011). *A human factors framework and study of the effect of nursing workload on patient safety and employee quality of working life*. *BMJ Quality & Safety*, 20(1): 15-24. <https://doi.org/10.1136/BMJQS.2008.028381>
- [10] Gumasing, M.J.J., Ilo, C.K.K. (2023). The impact of job satisfaction on creating a sustainable workplace: An empirical analysis of organizational commitment and lifestyle behavior. *Sustainability (Switzerland)*, 15(13): 10283. <https://doi.org/10.3390/su151310283>
- [11] Lu, Y., Hu, X.M., Huang, X.L., Zhuang, X.D., Guo, P., Feng, L.F., Hu, W., Chen, L., Zou, H., Hao, Y.T. (2017). The relationship between job satisfaction, work stress, work-family conflict, and turnover intention among physicians in Guangdong, China: A cross-sectional study. *BMJ Open*, 7(5): 1-12. <https://doi.org/10.1136/bmjopen-2016-014894>
- [12] Gazi, M.A.I., Yusof, M.F., Islam, M.A., Amin, M.B., bin S Senathirajah, A.R. (2024). Analyzing the impact of employee job satisfaction on their job behavior in the industrial setting: An analysis from the perspective of job performance. *Journal of Open Innovation: Technology, Market, and Complexity*, 10(4): 100427. <https://doi.org/10.1016/j.joitmc.2024.100427>
- [13] Riyanto, S., Handiman, U.T., Gultom, M., Gunawan, A., Putra, J.M., Budiyo, H. (2023). Increasing job satisfaction, organizational commitment and the requirement for competence and training. *Emerging Science Journal*, 7(2): 520-537. <https://doi.org/10.28991/ESJ-2023-07-02-016>
- [14] Thi Nong, N.M., Phuong, N.Q., Duc-Son, H. (2024). The effect of employee competence and competence – job – fit on business performance through moderating role of social exchange: A study in logistics firms. *Asian Journal of Shipping and Logistics*, 40: 187-197. <https://doi.org/10.1016/j.ajsl.2024.10.001>
- [15] Knight, C., Haslam, S.A. (2010). Your place or mine? Organizational identification and comfort as mediators of relationships between the managerial control of workspace and employees' satisfaction and well-being. *British Journal of Management*, 21(3): 717-735. <https://doi.org/10.1111/j.1467-8551.2009.00683.x>
- [16] Rahimnia, F., Eslami, G., Nosrati, S. (2019). Investigating the mediating role of job embeddedness: Evidence of Iranian context. *Personnel Review*, 48(3): 614-630. <https://doi.org/10.1108/PR-11-2017-0348>
- [17] Chernyak-Hai, L., Rabenu, E. (2018). The new era workplace relationships: Is social exchange theory still relevant? *Industrial and Organizational Psychology*, 11(3): 456-481. <https://doi.org/10.1017/iop.2018.5>
- [18] Lee, W., Migliaccio, G.C. (2018). Temporal effect of construction workforce physical strain on diminishing marginal productivity at the task level. *Journal of Construction Engineering and Management*, 144(9). [https://doi.org/10.1061/\(asce\)co.1943-7862.0001531](https://doi.org/10.1061/(asce)co.1943-7862.0001531)
- [19] Rojanasart, S., Edwards, N., Bhattacharyya, S. (2022). OP6 measuring and quantifying lost productivity due to priority diseases: A review of the past 10 years and reflections on the unprecedented changes and implications emerging with COVID-19. *Value in Health*, 25(7): S539. <https://doi.org/10.1016/j.jval.2022.04.1316>
- [20] Kim, S.J., Bae, S.J., Jang, M.W. (2022). Linear regression machine learning algorithms for estimating reference evapotranspiration using limited climate data. *Sustainability (Switzerland)*, 14(18): 11674. <https://doi.org/10.3390/su141811674>
- [21] Xu, C., Kong, Y. (2024). Random forest model in tax risk identification of real estate enterprise income tax. *PLoS One*, 19: 1-16. <https://doi.org/10.1371/journal.pone.0300928>
- [22] Ji, T., Wang, L., Yang, Y., Xu, L. (2023). Applied mathematics and nonlinear sciences application of random forest algorithm in the detection of foreign objects in wine. *Applied Mathematics and Nonlinear Sciences*, 9(1): 1-16. <https://doi.org/10.2478/10.2478/amns.2023.2.00055>
- [23] Uddin, N., Purwanto, E., Nugraha, H. (2024). Machine learning based modeling for estimating solar power generation. *E3S Web of Conferences*, 475: 1-12. <https://doi.org/10.1051/e3sconf/202447503009>
- [24] Uddin, N., Jaya, S., Purwanto, E., Putra, A.A.D., Fadhilah, M.W., Ramadhan, A.L.R. (2022). Machine-learning prediction of informatics students interest to the MBKM Program: A study case in Universitas Pembangunan Jaya. In *2021 International Seminar on Machine Learning, Optimization, and Data Science, ISMODE 2021*, pp. 146-151. <https://doi.org/10.1109/ISMODE53584.2022.9743125>
- [25] Humphrey, A., Cunha, P.A.C., Paulino-Afonso, A., Amarantidis, S., Carvajal, R., Gomes, J.M., Matute, I., Papaderos, P. (2023). Improving machine learning-derived photometric redshifts and physical property estimates using unlabelled observations. *Monthly Notices of the Royal Astronomical Society*, 520(1): 305-313. <https://doi.org/10.1093/mnras/stac3596>
- [26] Pereira, R., Junior, W., Guerra Junior, A.A. (2024). The

- use of Supervised Learning to perform pairwise classification for Record Linkage over real world data. *International Journal of Population Data Science*, 9(5 SE-Conference Proceedings).
<https://doi.org/10.23889/ijpds.v9i5.2902>
- [27] Valkenborg, D., Geubbelmans, M., Rousseau, A.J., Burzykowski, T. (2023). Supervised learning. *American Journal of Orthodontics and Dentofacial Orthopedics*, 164(1): 146-149.
<https://doi.org/10.1016/j.ajodo.2023.04.010>
- [28] Huang, S., Ailer, E., Kilbertus, N., Pfister, N. (2023). Supervised learning and model analysis with compositional data. *PLoS Computational Biology*, 19: 1-19. <https://doi.org/10.1371/journal.pcbi.1011240>
- [29] Ratkovic, M. (2023). Relaxing assumptions, improving inference: Integrating machine learning and the linear regression. *American Political Science Review*, 117(3): 1053-1069.
<https://doi.org/10.1017/S0003055422001022>
- [30] Cavalheiro, L. P., Bernard, S., Barddal, J.P., Heutte, L. (2024). Random forest kernel for high-dimension low sample size classification. *Statistics and Computing*, 34(1): 9. <https://doi.org/10.1007/s11222-023-10309-0>
- [31] Ismail, M.B.M., Mubarak, K.M., Azhar, A.G.M. (2015). Relationship between employee aspects and employee performance. *Journal of Management*, 12(1): 58-68.
- [32] Boström, M., Dellve, L., Thomée, S., Hagberg, M. (2008). Risk factors for generally reduced productivity—A prospective cohort study of young adults with neck or upper-extremity musculoskeletal symptoms. *Scandinavian Journal of Work, Environment & Health*, 34(2): 120-132.