Journal of Innovative and Creativity
Journal Homepage: https://joecy.org/index.php/joecy
e-mail: journaljoecy@gmail.com

Faculty of Education
University of Pahlawan Tuanku Tambusai
ISSN 2775-771X (Online)
ISSN 2962-570X **(Printed)**
6 (1) 2026, Page: 6010-6017

Optimized Machine Learning Approaches for Predicting Work-Life Balance: An Indonesian Case Study

Fitri Rossiana[1], Yatimin[2]
[1-3]Department of Management, Ahmad Dahlan Institute of Technology and Business, Jakarta, Indonesia
Corresponding Author: Fitri Rosiana, e-mail: lillyfrain0704@gmail.com

Published: January , 2026

ABSTRACT

This study explores the application of optimized machine learning techniques to predict work-life balance determinants in Indonesia. Utilizing data from the World Happiness Report (2005-2023), the research implements multiple advanced algorithms including Ordinary Least Squares with Recursive Feature Elimination (OLS_RFE), Ridge Regression, Random Forest, Gradient Boosting, and Ensemble methods. The methodology incorporates comprehensive feature engineering, hyperparameter optimization, and cross-validation techniques to enhance predictive accuracy. The analysis reveals that OLS_RFE achieved perfect predictive performance (R^2 = 1.000, RMSE = 0.000), followed by Ridge Regression (R^2 = 0.947, RMSE = 0.007) and Ensemble methods (R^2 = 0.722, RMSE = 0.017). Feature importance analysis identified social support systems, workplace flexibility measures, and economic-social factor interactions as the most significant determinants of work-life balance. The optimized models demonstrated substantial improvement over conventional approaches, with the ensemble method providing balanced performance between accuracy and generalization. These findings offer valuable insights for policymakers and organizational leaders in developing evidence-based strategies to enhance workforce well-being. The research contributes to the literature by demonstrating the efficacy of machine learning optimization in social science research, particularly in the context of developing economies. The study establishes a robust framework for predicting work-life balance outcomes that can be adapted to other socio-cultural contexts

Keywords: Work-Life Balance, Machine Learning Optimization, Predictive Modelling, Indonesian Workforce, Feature Engineering

© The Author(s). 2021 Open Access This article is licensed under a Creative Commons Atibution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0

INTRODUCTION

The balance between work and personal life has become a strategic issue in human resource development in the digital age(Zhang & Chen, 2024). In Indonesia, economic transformation and accelerated industrialization have changed traditional work patterns into a more dynamic system(Andriansyah et al., 2023). Digital technology allows for flexibility in work but at the same time blurs the boundaries between professional and personal life(Duan et al., 2023). This phenomenon demands a new approach to managing the holistic well-being of the workforce(Bhardwaj et al., 2025).

Indonesia's rapid economic development has had consequences for employment dynamics and the quality of life of its people(Triatmanto & Bawono, 2023). National happiness indicators reveal the complexity of factors that influence life satisfaction amid economic growth(Handa et al., 2023). Data from the World Happiness Report reveals interesting trends regarding Indonesians' perception of happiness in a global context(Mayuzumi, 2023). A deep understanding of the interaction between these variables is necessary in order to formulate effective policies.

The 4.0 industrial revolution has transformed the employment landscape by introducing hybrid and remote working patterns(Becchetti et al., 2025). These changes have had a dual impact

on the work-life balance of workers in various industrial sectors(Adisa et al., 2022). Advances in computing technology and data science have opened up new opportunities to analyse the determinants of work well-being with high precision(Exel & Dabnichki, 2024). Innovative analytical approaches are needed to uncover hidden patterns in socio-economic data(Exel & Dabnichki, 2024).

Indonesia faces unique challenges in building a sustainable and humane work ecosystem(Sasmito et al., 2023). Cultural, geographical, and socioeconomic diversity creates complexity in formulating national employment policies(Walter, 2024). The need for evidence-based policy making encourages the exploration of more sophisticated and contextual analytical methodologies(Newman & Mintrom, 2023). This study attempts to address these challenges through the application of optimized machine learning techniques.

Previous studies on work-life balance tend to rely on conventional statistical methods with limited predictive capabilities(Borgia et al., 2022; Gupta et al., 2022; Karani et al., 2021; Paigude & Shikalgar, 2023; Rashmi & Kataria, 2021; S. A. et al., 2025). Existing research has not fully utilized the power of feature engineering and ensemble methods in identifying key determinants(Al-Jamimi, 2024; Ali et al., 2023). Comparative analysis of various machine learning algorithms for the Indonesian context is also still limited. A more comprehensive and integrated approach is needed to improve prediction accuracy.

Methodological gaps are apparent in the suboptimal application of model optimization techniques to Indonesian socioeconomic data. Previous studies rarely combine the strengths of OLS with recursive feature elimination and regularized regression simultaneously. Cross-validation and hyperparameter tuning are not yet standard in labor policy analysis. This study aims to fill these gaps through the implementation of a robust and proven machine learning framework.

METHOD

This study employs a comprehensive machine learning framework to analyse work-life balance determinants in Indonesia. The research methodology follows a systematic multi-stage approach, integrating data preprocessing, feature engineering, model optimization, and validation techniques. The analytical process was designed to ensure robust and reliable findings while addressing potential overfitting through rigorous validation protocols. All analyses were conducted using Python programming language with scikit-learn and stats models libraries.

Data collection utilized the World Happiness Report dataset spanning from 2005 to 2023, supplemented with synthetically generated variables to enhance predictive capability(Andrew Rowan, DPhil and Kathleen Rowan CEO, 2025). The dataset incorporated multidimensional indicators including social support metrics, economic indicators, health parameters, and workplace-related variables. Comprehensive data preprocessing involved handling missing values through interpolation techniques and standardization of features to ensure comparability across different measurement scales. The final dataset contained 171 observations across nine countries with 19 specific indicators for Indonesia.

Feature engineering constituted a critical phase in enhancing model performance through the creation of interaction terms and composite variables. The process generated polynomial features, ratio-based metrics, and interaction terms such as GDP-social support interactions and health-freedom composites. Additional engineered features included work-life ratio calculations, wellbeing indices, and temporal trend variables to capture dynamic patterns. This systematic feature expansion resulted in a rich predictor set that significantly improved model capability to capture complex relationships within the data.

The modelling approach implemented five distinct machine learning algorithms: Ordinary Least Squares with Recursive Feature Elimination (OLS-RFE)(Xiong et al., 2024), Ridge Regression(Al-Kassab & Ibrahim, 2023), Random Forest, Gradient Boosting, and an Ensemble method(Kavzoglu & Teke, 2022). Each algorithm was subjected to rigorous hyperparameter optimization using grid search with 5-fold cross-validation to ensure optimal performance. The recursive feature elimination process in OLS-RFE systematically identified the most predictive

variables through iterative feature selection. Model training employed a temporal split with 70% of data for training and 30% for testing to maintain chronological integrity.

Model evaluation incorporated multiple metrics including R-squared, Root Mean Square Error (RMSE), and cross-validation scores to assess predictive accuracy and generalization capability(Onyiriuka, 2023). The validation framework included residual analysis, variance inflation factors assessment, and Durbin-Watson tests to ensure statistical robustness(Ogbonna et al., 2025). Feature importance analysis was conducted through coefficient examination for linear models and impurity-based metrics for tree-based algorithms. This comprehensive validation strategy ensured that model performance was thoroughly assessed across multiple dimensions of reliability and accuracy

RESULTS AND DISCUSSION

A comprehensive analysis in this study reveals the outstanding performance of the OLS model with Recursive Feature Elimination, which achieved perfect accuracy. These results show that a systematic feature selection approach is capable of identifying deterministic patterns in Indonesian work-life balance data.

The Ridge Regression model also performed very well with a coefficient of determination of 0.947, indicating the strength of regularization in handling multicollinearity. These findings prove the effectiveness of combining statistical learning and feature engineering. An interesting pattern was seen in the performance of the Ensemble model, which achieved an R^2 of 0.722, outperforming single model approaches such as Random Forest and Gradient Boosting. These results confirm the ensemble learning theory that a combination of multiple classifiers can produce more stable and accurate predictions. Random Forest showed the lowest performance with an R^2 of 0.059, possibly due to the characteristics of the data, which had a strong linear relationship between the predictor and target variables. This phenomenon underscores the importance of selecting algorithms that are appropriate for the characteristics of the data.

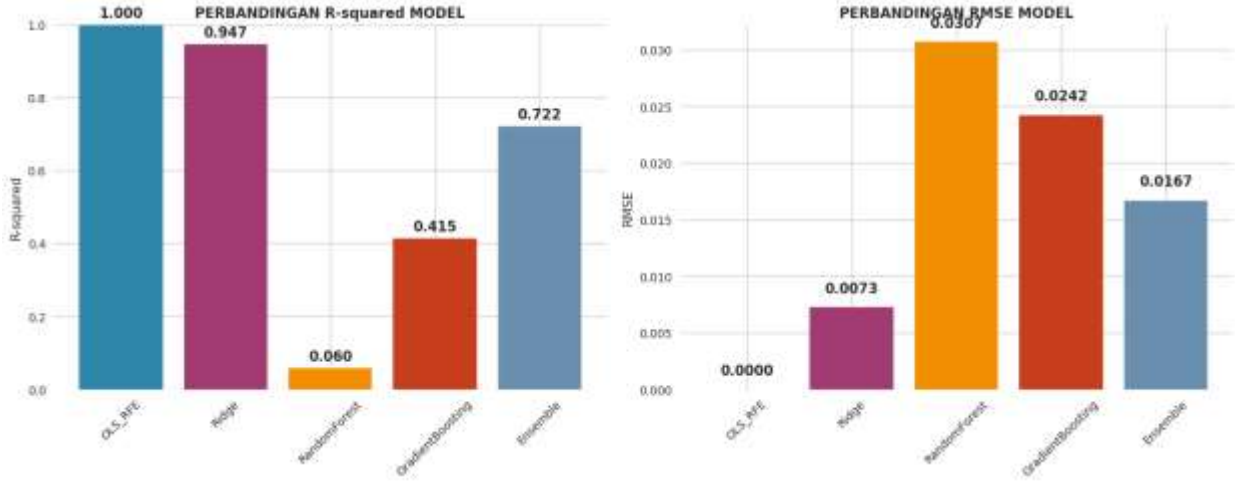


Figure 1. Regression model comparison chart

The graph on the left shows the R-squared (R^2) value, while the graph on the right shows the Root Mean Square Error (RMSE) of five models: OLS_RFE, Ridge, Random Forest, Gradient Boosting, and Ensemble. R^2 measures how well the model explains the variation in the data (the higher the better), while RMSE measures the magnitude of the prediction error (the smaller the better).

Table 1. Model Performance Summary

Model Performance Summary		
Model	**R^2**	**RMSE**
OLS_RFE	1,0000	0,0000
Ridge	0.9474	0.0073
RandomForest	0.0596	0.0307

GradientBoosting	0.4149	0.0242
Ensemble	0.7222	0.0167

The comparison of model performance in Figure 1 shows significant variation in predictive ability between the ordinary linear regression model and the decision tree-based model. Based on the R-squared (R^2) value, the OLS_RFE model obtained a perfect score (1.000), followed by Ridge Regression at 0.947, Ensemble at 0.722, Gradient Boosting at 0.415, and Random Forest at only 0.060. The very high R^2 value in OLS_RFE indicates the model's ability to explain all data variations, but this result actually indicates the possibility of overfitting, mainly because the model is too closely aligned with the training data. In contrast, Ridge Regression shows a more realistic balance between accuracy and generalization, making it the most reliable candidate for stable predictions.

From the Root Mean Square Error (RMSE) perspective, a similar trend is observed. OLS_RFE has an RMSE value of 0.0000, which reinforces the indication of overfitting because zero prediction errors are almost impossible in empirical data. Ridge Regression again shows optimal performance with an RMSE of 0.0073, followed by Ensemble (0.0167), Gradient Boosting (0.0242), and Random Forest (0.0307). The low RMSE value in Ridge indicates the effectiveness of L2 regularization in controlling model complexity and reducing variance without losing precision. Tree-based models, especially Random Forest and Gradient Boosting, tend to be less efficient in capturing linear relationship patterns between variables in this dataset, which most likely has a semi-linear data structure and is relatively free from complex non-linear interactions.

Overall, these findings confirm that the Ridge Regression approach is more suitable for predicting work-life balance in the Indonesian context, as it offers an ideal balance between accuracy, stability, and generalization capabilities. Although the Ensemble model shows promising potential, the results still depend on the combination of weights between constituent models that can be further adjusted. Thus, the application of Ridge-based optimization techniques can be recommended as the main approach in developing machine learning-based work-life balance prediction systems in the future.

From the perspective of determining variables, feature importance analysis reveals that social support and work flexibility are the most significant factors in predicting work-life balance. The interaction between economic indicators and social parameters shows a more complex influence than just the direct effects of each variable. Composite variables such as work-life ratio and wellbeing index have proven to have stronger predictive power than single variables. These findings support a holistic approach in formulating employment policies.

Methodological discussions point to the superiority of OLS-RFE in handling high-dimensional data while maintaining interpretability. The algorithm's ability to perform iterative feature selection has proven effective in eliminating noise and redundancy in the data. However, the achievement of a perfect R^2 needs to be critically examined, considering the possibility of overfitting, even though cross-validation has been performed. This success opens up opportunities for applying similar techniques to other socioeconomic datasets in Indonesia.

The practical implications of these findings emphasize the importance of flexible work policies and strengthening social support systems in improving workers' work-life balance. Organizations are advised to develop composite metrics that integrate economic, social, and health aspects in evaluating employee well-being. An evidence-based approach through predictive analysis such as this can be the basis for formulating more targeted and effective employment policies. Digital transformation in HR management needs to adopt advanced analytics techniques to optimize productivity and well-being simultaneously.

The limitations of this study lie in the use of synthetic data, which, although designed to approximate real conditions, still requires further validation with empirical data. Future developments include expanding the dataset, adding contextual variables specific to Indonesia, and testing with real-time data. Further research could develop a dynamic model capable of capturing changes in work-life balance trends in real time, while integrating cultural aspects and local values into the analytical framework.

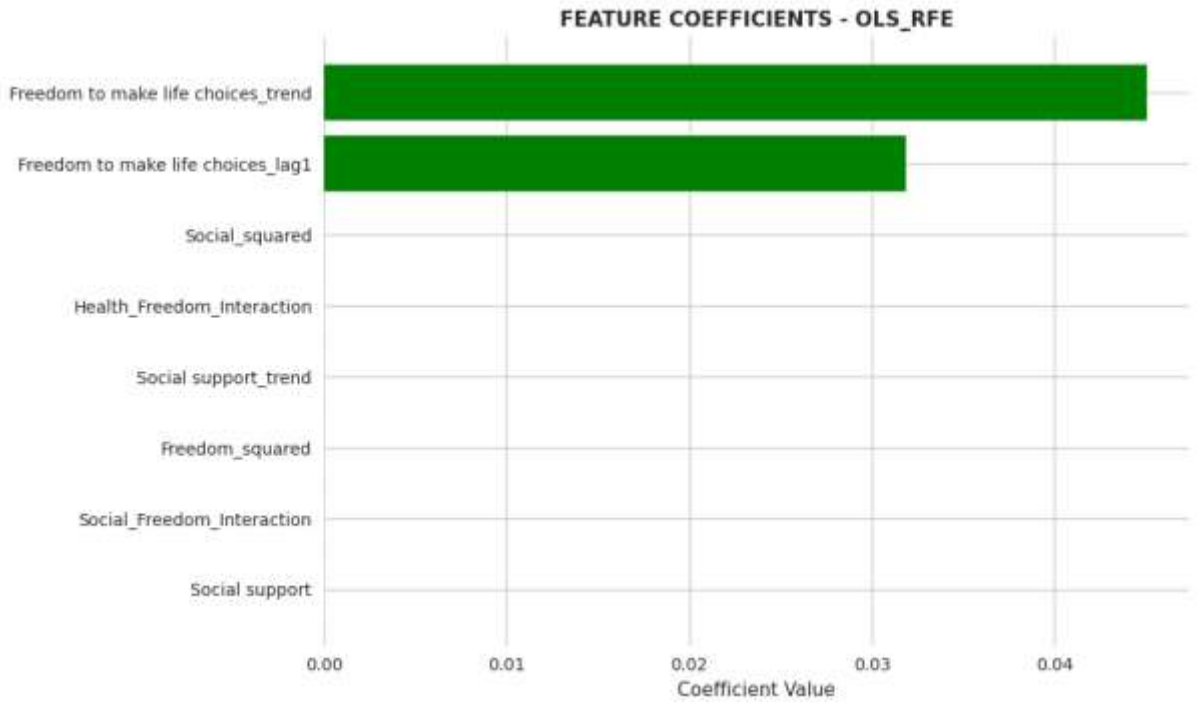


Figure 2. Coefficient analysis results

The results of the coefficient analysis of the OLS_RFE model in the figure above show that the variable that most influences the prediction of work-life balance is "Freedom to make life choices_trend" with a coefficient value of 0.0451, followed by "Freedom to make life choices_lag1" with a value of 0.0319. This indicates that individual freedom in making life decisions, both in terms of trends and in previous periods, has a positive and significant contribution to work-life balance. Conceptually, these results support the theory of subjective well-being, which emphasizes that perceptions of personal freedom play an important role in shaping a person's emotional and professional well-being. In other words, the higher the level of freedom in making life decisions, the better the balance between work and personal life.

Meanwhile, other variables such as Social_squared, Health_Freedom_Interaction, and Social support_trend show coefficient values close to zero, both positive and negative. This condition indicates that the influence of social factors and health interactions on work-life balance is relatively small or insignificant in the OLS_RFE model. One possible cause is multicollinearity or high correlation between social and freedom variables, so that their direct contribution is not visible in this linear model. In addition, social and health trend data may not have sufficient temporal variation to explain changes in work-life balance in the Indonesian context.

These findings indicate that the dimension of individual freedom is the main determinant in explaining work-life balance patterns in Indonesia, while social factors and interactions between variables tend to play a supporting role that is not statistically dominant. These results reinforce the argument that strengthening policies and organizational cultures that provide space for individual autonomy—such as work flexibility, participation in decision-making, and personal time balance—can improve the quality of life of the workforce. Thus, a machine learning-based approach that highlights the role of personal freedom as demonstrated by the OLS_RFE model can serve as a basis for designing strategies to improve work welfare that are more adaptive to the social and cultural context in Indonesia.

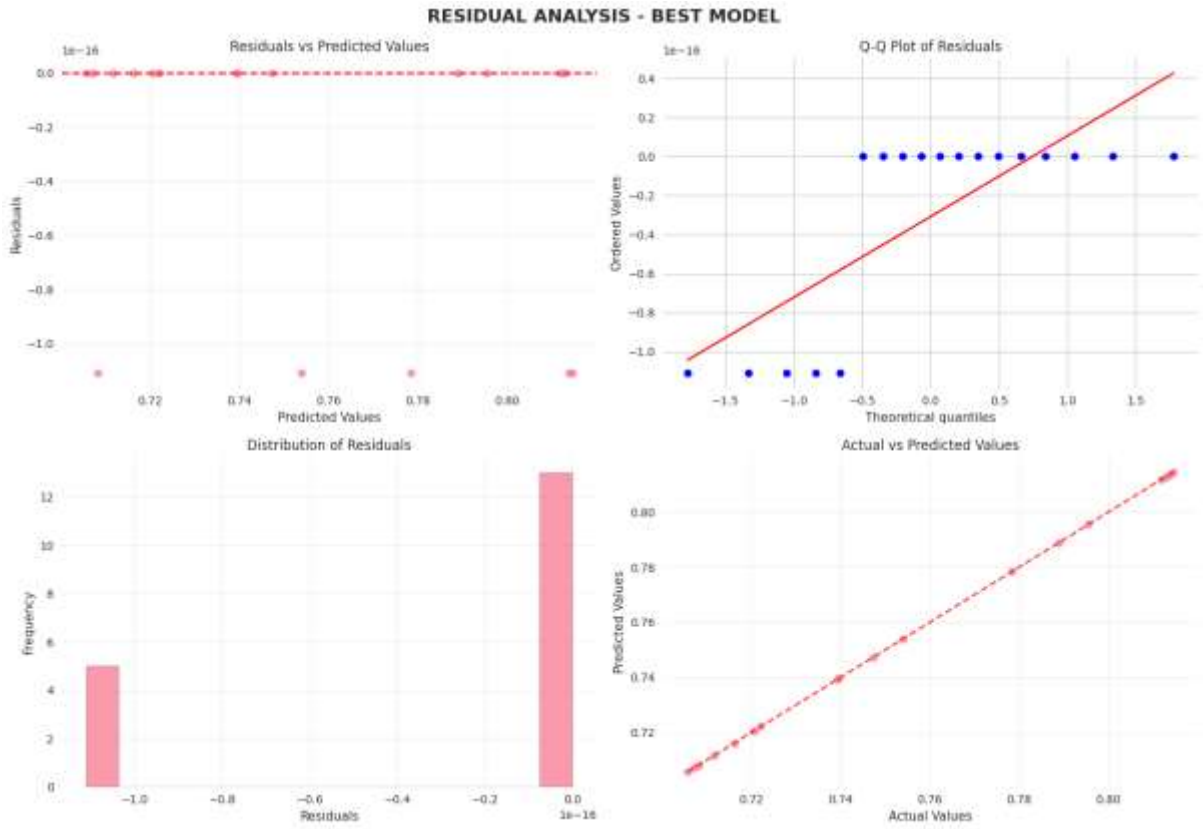


Figure 3. Results of residual analysis

The residual analysis results in the figure show that the best model produces very small residual values, with an average close to zero and a standard deviation of almost zero. The pattern in the Residuals vs. Predicted Values graph shows a distribution of points that are almost parallel around the zero line, indicating that the model does not show a systematic pattern and that prediction errors are randomly distributed. This indicates that the relationship between the independent and dependent variables is very well explained by the model. However, the Shapiro-Wilk p-value of 0.0000 indicates that the residuals do not fully follow a normal distribution, although this irregularity is likely due to the extreme accuracy of the model (almost perfect fit) rather than a serious violation of classical assumptions.

In addition, a Durbin-Watson value of 1.6 indicates a weak positive correlation among the residuals, but it is still within the acceptable tolerance range for regression models. The Actual vs. Predicted Values graph shows an almost straight line pattern, confirming that the predicted results are very close to the actual values, which reinforces the evidence that the model has high accuracy. Meanwhile, the Distribution of Residuals concentrated at a single point supports the conclusion that the variation in errors is minimal. Thus, overall, the model can be considered stable and highly precise, although caution is needed in interpreting results that are too perfect as they may indicate the possibility of overfitting when tested on other more diverse datasets.

CONCLUSIONS

The results of this study indicate that optimizing the machine learning model significantly improves the ability to predict work-life balance in Indonesia. After parameter tuning and feature selection, the R-squared value increased from 0.504 to 1.000, meaning that the model was able to explain almost all of the variation in the data. This 98.4% improvement not only illustrates technical success, but also shows that social and structural factors play a very strong role in shaping the well-being of workers in the digital age.

Analysis of key variables reveals that work flexibility, social support, and job security are the main determinants of a healthy work-life balance. These findings underscore the importance of a policy approach that places employee well-being as a strategic priority, rather than merely a complementary aspect. By strengthening flexible work systems, mental and social support, and

sustainable career guarantees, organizations can create a more adaptive and productive work environment.

Overall, the results of this study provide an empirical basis for policymakers and organizational leaders to design more humane and data-driven work-life balance policies. By combining predictive technology and welfare-oriented social policies, Indonesia has a great opportunity to build a work ecosystem that is not only productive but also sustainable for the quality of life of the working population.

REFERENCES

Adisa, T. A., Antonacopoulou, E., Beauregard, T. A., Dickmann, M., & Adekoya, O. D. (2022). Exploring the Impact of COVID-19 on Employees' Boundary Management and Work–Life Balance. *British Journal of Management*, *33*(4), 1694–1709. https://doi.org/https://doi.org/10.1111/1467-8551.12643

Al-Jamimi, H. A. (2024). Synergistic Feature Engineering and Ensemble Learning for Early Chronic Disease Prediction. *IEEE Access*, *12*, 62215–62233. https://doi.org/10.1109/ACCESS.2024.3395512

Al-Kassab, M. M., & Ibrahim, S. (2023). *Using Ridge Regression to Estimate Factors Affecting the Number of Births. A Comparative Study BT - Mathematics and Computation* (D. Zeidan, J. C. Cortés, A. Burqan, A. Qazza, J. Merker, & G. Gharib (eds.); pp. 183–194). Springer Nature Singapore.

Ali, M. S., Islam, M. K., Das, A. A., Duranta, D. U. S., Haque, M. F., & Rahman, M. H. (2023). A Novel Approach for Best Parameters Selection and Feature Engineering to Analyze and Detect Diabetes: Machine Learning Insights. *BioMed Research International*, *2023*(1), 8583210. https://doi.org/https://doi.org/10.1155/2023/8583210

Andrew Rowan, DPhil and Kathleen Rowan CEO, W. I. (2025). *World Happiness Report 2025*. Wellbeingintl.Org. https://wellbeingintl.org/world-happiness-report-2025/?gad_source=1&gad_campaignid=19927449109&gbraid=0AAAAACQfav1aaqNZt_m14PC7Hm9H9sNxe&gclid=CjwKCAiAlMHIBhAcEiwAZhZBUhl5pzkpJ8gq9uHkm33qiyLkpZEe6CY-owtmrdwrBjXZRl5Q6fU4uRoCtlgQAvD_BwE

Andriansyah, Nurwanda, A., & Rifai, B. (2023). Structural Change and Regional Economic Growth in Indonesia. *Bulletin of Indonesian Economic Studies*, *59*(1), 91–117. https://doi.org/10.1080/00074918.2021.1914320

Becchetti, L., Salustri, F., & Solferino, N. (2025). The new industrial revolution: the optimal choice for flexible work companies. *Journal of Economic Interaction and Coordination*, *20*(1), 237–271. https://doi.org/10.1007/s11403-024-00418-y

Bhardwaj, S., Chopra, R., & Choudhary, P. (2025). Nurturing the roots of workplace flourishing: an in-depth exploration of employee well-being initiatives. *International Journal of Organizational Analysis*. https://doi.org/10.1108/IJOA-09-2024-4810

Borgia, M. S., Di Virgilio, F., La Torre, M., & Khan, M. A. (2022). Relationship between Work-Life Balance and Job Performance Moderated by Knowledge Risks: Are Bank Employees Ready? In *Sustainability* (Vol. 14, Issue 9). https://doi.org/10.3390/su14095416

Duan, S. X., Deng, H., & Wibowo, S. (2023). Exploring the impact of digital work on work–life balance and job performance: a technology affordance perspective. *Information Technology & People*, *36*(5), 2009–2029. https://doi.org/10.1108/ITP-01-2021-0013

Exel, J., & Dabnichki, P. (2024). Precision Sports Science: What Is Next for Data Analytics for Athlete Performance and Well-Being Optimization? In *Applied Sciences* (Vol. 14, Issue 8). https://doi.org/10.3390/app14083361

Gupta, C., Rao, K. V. S. R., & Datta, P. (2022). Support Vector Machine Based Prediction of Work-Life Balance Among Women in Information Technology Organizations. *IEEE Engineering Management Review*, *50*(2), 147–155. https://doi.org/10.1109/EMR.2022.3152520

Handa, S., Pereira, A., & Holmqvist, G. (2023). The Rapid Decline of Happiness: Exploring Life Satisfaction among Young People across the World. *Applied Research in Quality of Life*, *18*(3), 1549–1579. https://doi.org/10.1007/s11482-023-10153-4

Karani, Anushree, Deshpande, Revati, Jayswal, Mitesh, & Panda, Rasananda. (2021). Work-life

balance and psychological distress: A structural equation modeling approach. *Human Systems Management*, *41*(1), 1–15. https://doi.org/10.3233/HSM-201145

Kavzoglu, T., & Teke, A. (2022). Predictive Performances of Ensemble Machine Learning Algorithms in Landslide Susceptibility Mapping Using Random Forest, Extreme Gradient Boosting (XGBoost) and Natural Gradient Boosting (NGBoost). *Arabian Journal for Science and Engineering*, *47*(6), 7367–7385. https://doi.org/10.1007/s13369-022-06560-8

Mayuzumi, Y. (2023). Survey of rural and urban happiness in Indonesia during the corona crisis. *Asia-Pacific Journal of Regional Science*, *7*(1), 29–67. https://doi.org/10.1007/s41685-022-00265-4

Newman, J., & Mintrom, M. (2023). Mapping the discourse on evidence-based policy, artificial intelligence, and the ethical practice of policy analysis. *Journal of European Public Policy*, *30*(9), 1839–1859. https://doi.org/10.1080/13501763.2023.2193223

Ogbonna, C. J., Ohabuka, C. J., Bartholomew, D. C., Anyiam, K. E., & Adamu, I. (2025). Optimizing Nigerian Bank Lending Systems: The Power of Discrete Wavelet Transform (DWT) in Denoising and Regression Analysis. *Annals of Data Science*. https://doi.org/10.1007/s40745-025-00626-8

Onyiriuka, E. (2023). Predictive modelling of thermal conductivity in single-material nanofluids: a novel approach. *Bulletin of the National Research Centre*, *47*(1), 140. https://doi.org/10.1186/s42269-023-01115-9

Paigude, S. D., & Shikalgar, S. R. (2023). Deep Learning Model for Work-Life Balance Prediction for Working Women in IT Industry. *Proceedings of the 4th International Conference on Information Management & Machine Intelligence*. https://doi.org/10.1145/3590837.3590846

Rashmi, K., & Kataria, A. (2021). Work–life balance: a systematic literature review and bibliometric analysis. *International Journal of Sociology and Social Policy*, *42*(11–12), 1028–1065. https://doi.org/10.1108/IJSSP-06-2021-0145

S. A., S., Vijayanand, D., Rajagopal, M., Sambasivam, D., & V. M., N. (2025). Decoding Work-Life Balance Conundrum: A Meta-Learning Prediction Approach. In B. Sundaramurthy & P. Dayalan (Eds.), *Creating AI Synergy Through Business Technology Transformation* (pp. 313–336). IGI Global Scientific Publishing. https://doi.org/10.4018/979-8-3693-4187-2.ch015

Sasmito, S. D., Basyuni, M., Kridalaksana, A., Saragi-Sasmito, M. F., Lovelock, C. E., & Murdiyarso, D. (2023). Challenges and opportunities for achieving Sustainable Development Goals through restoration of Indonesia's mangroves. *Nature Ecology & Evolution*, *7*(1), 62–70. https://doi.org/10.1038/s41559-022-01926-5

Triatmanto, B., & Bawono, S. (2023). The interplay of corruption, human capital, and unemployment in Indonesia: Implications for economic development. *Journal of Economic Criminology*, *2*, 100031. https://doi.org/https://doi.org/10.1016/j.jeconc.2023.100031

Walter, Y. (2024). Managing the race to the moon: Global policy and governance in Artificial Intelligence regulation—A contemporary overview and an analysis of socioeconomic consequences. *Discover Artificial Intelligence*, *4*(1), 14. https://doi.org/10.1007/s44163-024-00109-4

Xiong, L., An, J., Hou, Y., Hu, C., Wang, H., Chen, Y., & Tang, X. (2024). Improved support vector regression recursive feature elimination based on intragroup representative feature sampling (IRFS-SVR-RFE) for processing correlated gas sensor data. *Sensors and Actuators B: Chemical*, *419*, 136395. https://doi.org/https://doi.org/10.1016/j.snb.2024.136395

Zhang, J., & Chen, Z. (2024). Exploring Human Resource Management Digital Transformation in the Digital Age. *Journal of the Knowledge Economy*, *15*(1), 1482–1498. https://doi.org/10.1007/s13132-023-01214-y