Analysis of Well-Being and Lifestyle Data: Predicting Work-Life Balance Scores for Men

Introduction

This technical paper explores the predictors of men's Work-Life Balance (WLB) using a machine learning approach based on lifestyle factors. The dataset, sourced from Kaggle, includes 24 variables that represent various lifestyle habits such as number of sleep hours, support for others, places visited, and one's social circle size.

Our goal was to develop a predictive model capable of accurately estimating WLB scores for men and identifying the most influential factors for a better work-life balance.

The importance of this research lies in the limited exploration of work-life balance, despite its growing relevance. In today's fast-paced world, the pressures of meeting increasing demands often disrupt this balance, leading to issues such as fatigue and mental health challenges. A personal example is managing the demands of a full-time, high-pressure job while simultaneously pursuing a rigorous AI master's program, highlighting the real-world implications of this research.

In a peer reviewed article "Methodological Choices in Work-Life Balance Research 1987 to 2006: A Critical Review" highlights the importance of methodological consistency in the field, noting that "work-life balance studies need to establish greater consistency between the conceptualization of constructs and the operationalization of measures" (Chang, McDonald, & Burton, 2009). This underscores the challenges researchers face in maintaining methodological rigor while studying complex social constructs like work-life balance.

Through advancing machine learning (ML) research in men's health, our goal is to uncover key habits that can support men in navigating the evolving and demanding commitments of modern life.

Data Preparation

The original dataset included both male and female respondents; however, for our analysis goal, we focused solely on male entries.

The data was pre-processed by:

- Removing non-integer variables, such as AGE, Timestamp, and GENDER, for compatibility in quantitative analysis.
- Handling missing values in the DAILY_STRESS column by converting non-numeric values and removing rows with null entries.

After pre-processing, the refined dataset—comprising 24 variables and 6,114 male respondents—was split into training and test sets using a random sampling with replacement approach.

Model Selection: Random Forest

To predict the WLB scores, a Random Forest Regressor was chosen due to its ability to handle both linear and non-linear relationships, and its inherent ability to prevent overfitting through ensemble learning. The model was trained on the lifestyle variables to predict the WLB score.

Model Performance

Key performance metrics of the model were:

- Mean Squared Error (MSE): **158.31**
- R-squared (R^2): **0.9266**

With an R^2 value of 0.93, the model explained 93% of the variance in WLB scores, demonstrating its robustness in capturing key patterns within the data. This high R^2 indicates that the model can accurately predict work-life balance (WLB) scores based on the selected lifestyle variables.

However, MSE of 158.31 suggests there is still some room for improvement, as reducing the MSE could lead to more precise predictions.

While the model performs well overall, further optimization—such as exploring additional features, may enhance accuracy and minimize MSE.

Feature Importance

The Random Forest model highlighted key lifestyle variables that significantly influence men's WLB scores.

a) **Achievement**: The most influential predictor, accounting for 26.3% of the variance. This indicates that a sense of accomplishment plays a central role in balancing work and personal life.
b) **Task Completion**: Explained 10.1% of the variance, emphasizing the importance of productivity and the satisfaction gained from completing daily tasks.
c) **Supporting Others**: Contributed 9.98% of the variance, underscoring the positive impact of providing social support on overall well-being.
d) **Places Visited**: Explained 7.88% of the variance, suggesting that exposure to new environments and experiences contributes meaningfully to a balanced lifestyle.
e) **Sufficient Income**: Accounted for 5.72% of the variance, reflecting that financial stability, though not the most dominant factor, still plays a significant role in maintaining work-life balance.

Above findings illustrate that personal achievements, productivity, social interactions, and financial security are pivotal in shaping men's overall well-being.

By understanding these relationships, men can make targeted adjustments to improve their work-life balance.

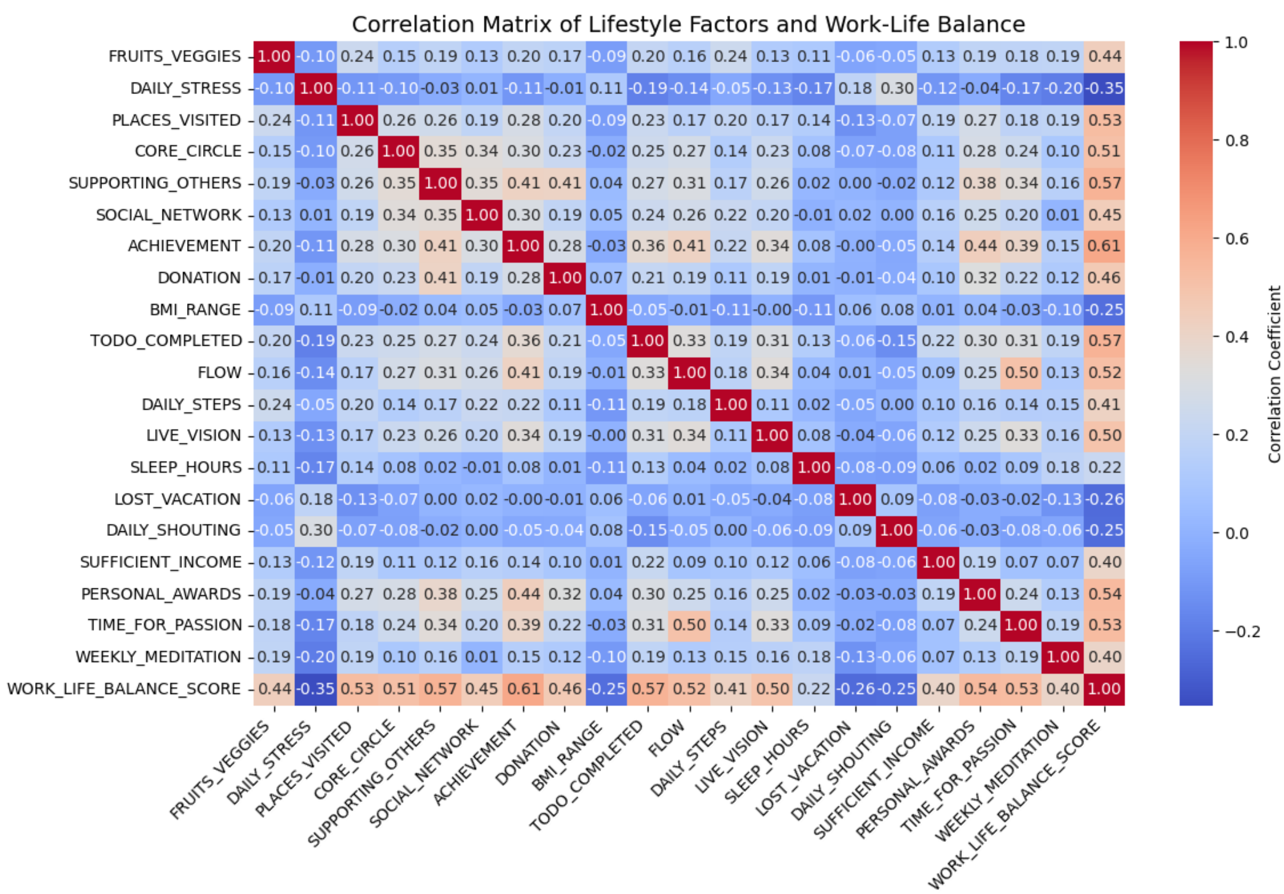
Alternative Exploration of Variables

To gain deeper insights into the relationships between lifestyle factors and WLB scores, we analyzed the variables using correlation matrix and boxplots.

Correlation Matrix

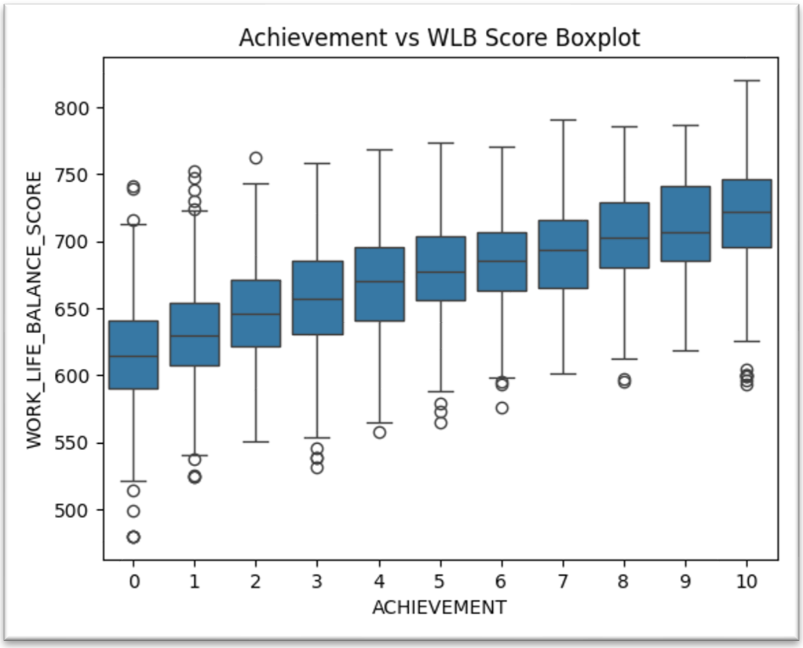
The correlation matrix revealed that **Supporting Others (0.53)** and **Task Completion (0.54)** have the strongest positive correlations with WLB score, indicating that social support and productivity play significant roles in well-being.

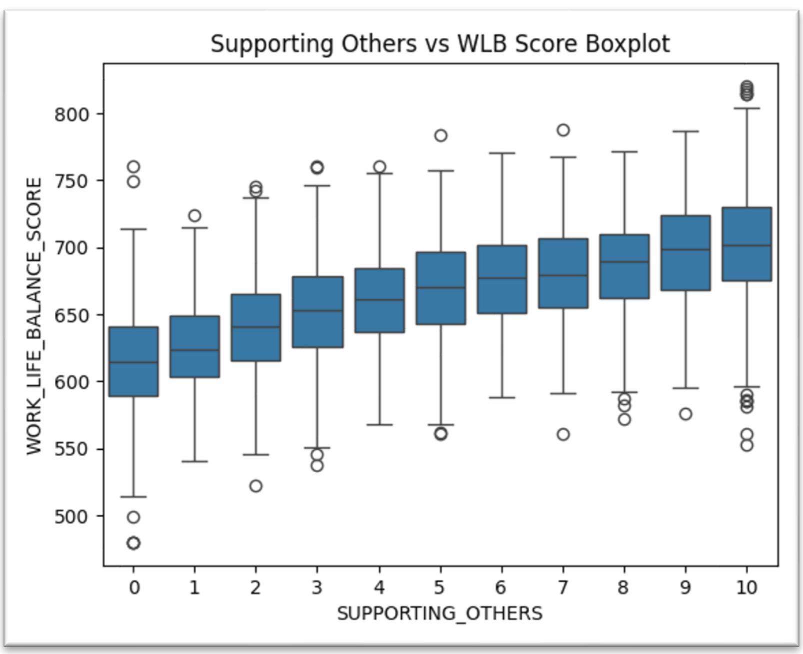
Achievement (0.44) also shows a notable positive correlation, highlighting the impact of personal success. Meanwhile, **Places Visited (0.28)** and **Sufficient Income (0.25)** display more modest correlations.



Boxplot Insights

The boxplot visualizations show that higher level of **Achievement** is consistently associated with better WLB scores, reinforcing its critical role in life balance. **Task Completion** (*visual omitted)* and **Supporting Others** also show strong positive relationships with WLB, suggesting that productivity and social engagement significantly enhance well-being.





Statistical Approach

Our analysis leverages the strength of Random Forest in capturing both linear and non-linear interactions between variables. The model's performance metrics, including the MSE of 158.31 and the R^2 of 0.93, underscore its accuracy.

These results are further supported by traditional linear methods, such as the correlation matrix, which identified direct linear relationships between variables like **Supporting Others** and **Task Completion**.

The boxplots offered a visual confirmation of the relationships identified by the model. This combined statistical and visual approach allowed us to identify which variables significantly contribute to WLB and *verify* the robustness of the Random Forest model.

Conclusion & Next Steps

Our Random Forest model was effective in identifying the most important predictors and generating accurate WLB predictions.

While the model performed well, there is potential for improvement.

Moving forward, there are opportunities to:

- Explore additional variables: Incorporating new features, such as **Flow** or **Time for Passion,** may further enhance the model.
- Refine existing variables: Expanding the limited range of **Sufficient Income** and addressing potential scaling inconsistencies in other variables could yield better results.
- Apply advanced techniques: Exploring techniques like Gradient Boosting or Neural Networks could improve performance, especially in capturing more complex relationships between variables.

Reference:

Chang, A., McDonald, P., & Burton, P. (2010). Methodological choices in work-life balance research 1987 to 2006: a critical review. *International Journal of Human Resource Management*, *21*(13), 2381–2413. https://doi.org/10.1080/09585192.2010.516592