Modeling the Link Between Sleep Behavior and Productivity: A Machine Learning Approach

Yiyang Bu, Zixin Fang, Keer Hu, Jessica Xiong, Han Yan
DS3001 – Foundations of Machine Learning
University of Virginia, May 2025

May 7, 2025

Abstract

This project explores whether individual sleep behaviors, lifestyle factors, and daily habits can help explain fluctuations in self-reported productivity. Using a dataset of 5,000 synthetic day-level records—including variables such as total sleep hours, sleep quality, stress level, caffeine intake, screen time before bed, and mood—we engineered features like sleep consistency and sleep efficiency to better capture behavioral patterns. We implemented a full machine learning pipeline, including standardization, LASSO feature selection, principal component analysis (PCA), and four predictive models: linear regression, decision trees, k-nearest neighbors (kNN), and artificial neural networks (ANN). While these features are commonly believed to impact daily performance, our findings show that sleep behavior alone is not sufficient to predict productivity. Across all models, R^2 scores were negative or close to zero, and complex models failed to outperform simpler baselines. Although PCA improved model stability, it limited interpretability of individual features. Overall, our results highlight the multi-dimensional nature of productivity and the limitations of using behavioral data alone to model such outcomes.

1 Introduction

Sleep is often emphasized as a cornerstone of mental health, focus, and overall well-being. With the rise of sleep-tracking apps and devices that promise to optimize rest, the assumption that "better sleep means better performance" has become widely accepted. But how strong is that connection—especially on a day-to-day basis? Does getting more sleep or going to bed earlier actually lead to higher productivity the next day?

This project focuses on a practical and testable question: can sleep-related behaviors help predict a person's next-day productivity? Rather than relying on general assumptions, we treat this as a data-driven problem. Our goal is to examine whether machine learning models can detect meaningful patterns in sleep and lifestyle data that relate to productivity. Instead of focusing only on sleep duration, we include features such as sleep quality, stress level, screen time before bed, and mood—variables that capture more nuanced aspects of daily behavior.

We used a dataset that contains 5,000 synthetic entries, each representing a single day in the life of an individual. Each observation combines behavioral metrics (like caffeine intake, work hours, and exercise) with self-reported mental states and a productivity score ranging from 1 to 10. The structure of the dataset allows us to explore daily variation while controlling for individual differences, though it also introduces potential issues with noise, subjectivity, and missing context.

To investigate, we used supervised regression models of varying complexity: linear regression, decision trees, kNN, and ANN. We applied LASSO regression to reduce noise, PCA to handle multicollinearity, and standardization to support distance-based models. All modeling decisions were grounded in tools and workflows from the DS4E GitHub repository to stay within the scope of the course.

Our findings suggest that while some weak patterns may exist, none of the models produced reliable predictions. In fact, no model meaningfully outperformed a baseline that simply predicted the average score. PCA helped improve model performance marginally, but made it more difficult to interpret which specific behaviors were most influential. These results suggest that if a relationship between sleep and productivity exists, it is likely weak, nonlinear, and shaped by unobserved external factors.

The remainder of the paper is organized as follows: Section 2 describes the dataset and variables. Section 3 details our preprocessing steps and modeling choices. Section 4 presents the results of our analysis. Section 5 offers a broader reflection on the challenges we encountered and outlines future directions.

2 Data

The dataset used in this project consists of 5,000 synthetic day-level records, each representing a single day in the life of one individual. These records include both objective and subjective variables, such as total sleep hours, caffeine intake, mood score, and a self-assessed productivity score ranging from 1 to 10.

To explore initial relationships in the data, we first examined sleep hour distributions and the association between sleep and productivity using visualizations.

As shown in Figure 1, productivity has weak linear correlation with most features, especially sleep-related variables like total sleep hours and mood score. These observations motivated our use of more flexible, non-linear models.

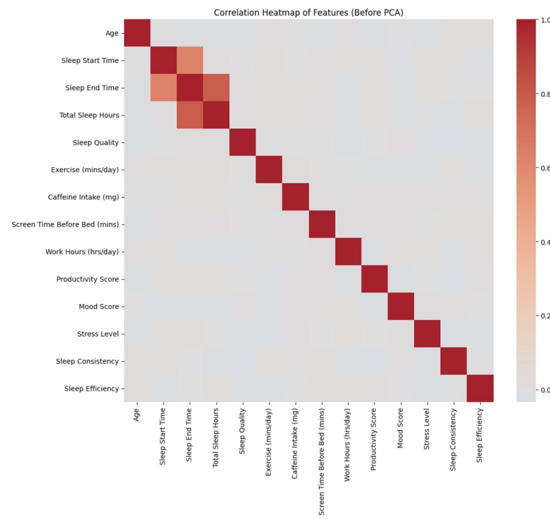

Figure 1: Correlation Heatmap of Numeric Features (Before PCA). Heatmap visualizing linear correlations between numeric features. Productivity shows little association with any individual feature.

3 Methods

3.1 Preprocessing

We engineered new features, including sleep consistency (standard deviation of sleep hours per individual), sleep efficiency (total sleep hours divided by the time between sleep start and end), and an evening screen ratio (screen time before bed normalized by sleep window length in minutes). We standardized numerical features and applied one-hot encoding to categorical variables as needed.

3.2 Modeling

We implemented four models: linear regression, decision tree, k-nearest neighbors (kNN), and artificial neural networks (ANN). To reduce overfitting and improve numerical stability, we used LASSO for feature selection and PCA to retain 95% of total variance.

3.3 Evaluation

Models were evaluated using RMSE, MAE, and R^2. These metrics provided insight into prediction accuracy, average error, and explained variance.

4 Results

Across all models, R^2 scores were consistently negative or near zero. Linear regression achieved the best relative performance with RMSE = 1.0165, MAE = 0.8832, and R^2 = -0.0092. More complex models failed to outperform this baseline. ANN models also showed signs of overfitting.

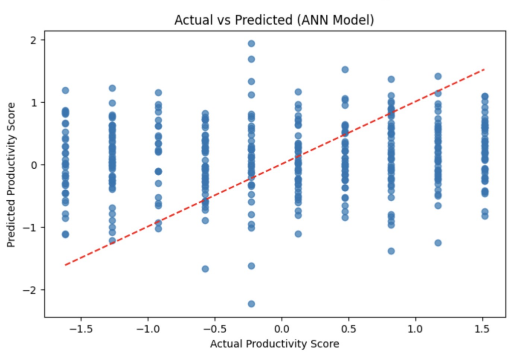

Figure 2: Predicted vs. Actual Productivity Scores (ANN Model). The wide spread reflects poor generalization.

Figure 2 shows predicted vs. actual productivity for the ANN model. The lack of clustering along the diagonal reflects poor model generalization.

Figure 3 displays the most important features from the decision tree. Because PCA was used, these are principal components rather than original variables, making interpretation limited.

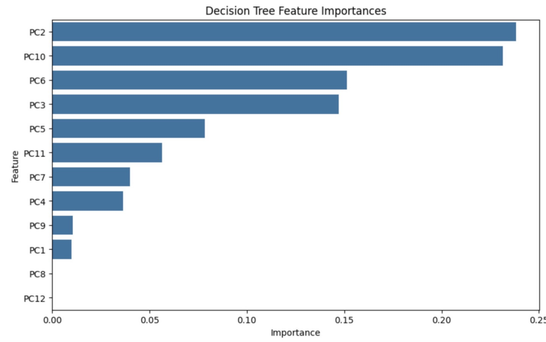

Figure 3: Feature Importance from Decision Tree Model. Most predictive features are PCA components, not interpretable raw variables.

5 Conclusion

This project set out to examine whether machine learning models can predict productivity based on sleep-related behaviors. While this idea is intuitive, our results show that behavioral data alone is not sufficient. Across models, R^2 values were weak or negative. We observed overfitting in ANN models and limited interpretability in PCA components.

These results reflect broader issues with modeling subjective outcomes like productivity. Key influences—like emotional resilience or task difficulty—were missing from the dataset. The use of self-reported productivity also introduced noise.

Future work could use more objective productivity measures and richer data sources, such as biometric tracking or time-logging tools. Despite weak predictive power, our findings reinforce the need for thoughtful feature engineering and critical evaluation of behavioral data in modeling complex outcomes.