

# DreamHarmony: Unveiling Sleep Patterns through Lifestyle Modeling and Predictive Analytics

Hoang Q. Nguyen<sup>\*†</sup>, Md Mahmudul Hassan<sup>\*†</sup>, Solomon Rukundo<sup>†‡</sup>

<sup>\*</sup> Korea Institute of Science and Technology

<sup>†</sup> Korea Research Institute of Standards and Science

<sup>‡</sup> Korea National University of Science and Technology (UST)

**Abstract**—This study investigates how daily habits like exercise and technology use are related to sleep quality, focusing on associated stress and heart health. After careful data preparation, we used statistical methods, including Pearson [1] and Spearman correlations [2], to identify which lifestyle factors are most connected to good or poor sleep. We then built various mathematical models, such as linear regression and decision trees, to predict sleep quality based on these factors. Our results show that some daily activities have a strong link to how well people sleep, and our models can explain a significant part of this link. These findings could help guide recommendations for improving sleep based on personal habits.

**Index Terms**—big data, sleep, health, lifestyle, machine learning, support vector machine

## I. INTRODUCTION

The interplay between lifestyle and sleep quality is an increasingly pertinent subject in the realm of public health. To explore this intricate relationship, we embarked on a comprehensive data collection campaign, developing a structured survey that captures a wide array of lifestyle variables alongside sleep metrics. The survey’s design allowed for the nuanced capture of daily routines, technological interactions, exercise habits, and their temporal association with sleep patterns.

Post-collection, data processing commenced with rigorous cleaning techniques to ensure integrity and analytical viability. Our initial exploratory analysis hinged on the construction of a correlation matrix heatmap, which served as a visual aid to discern potential relationships between variables. This graphical representation provided the preliminary clues necessary for subsequent hypothesis formulation.

Armed with hypotheses, we delved into a variety of statistical tests. Spearman’s rank correlation offered insights into monotonic relationships, while Pearson’s correlation assessed linear dependencies. For categorical data, the Chi-squared test [3] evaluated associations between discrete variables. These statistical methodologies were pivotal in distilling the factors most relevant to sleep quality.

The culmination of our hypothesis testing informed the selection of variables for model fitting. We implemented a gamut of machine learning models, including, but not limited

to, regression analyses, decision trees, and support vector machines. Each model was rigorously evaluated to determine its predictive power and relevance to our study.

Our contributions through this research are multifaceted. Initially, we developed a foundational survey meticulously engineered to assess sleep and lifestyle variables. Subsequently, we discerned pivotal factors that influence sleep quality by leveraging statistical analysis and hypothesis testing. Furthermore, we harnessed machine learning algorithms to construct models based on these factors. Our study establishes a framework for future exploration, highlighting the necessity for continuous data accumulation to refine and augment model efficacy. We invite researchers and the public alike to contribute to our ongoing work (available on GitHub<sup>1</sup>) through feedback and new applications of our findings.

## II. RELATED WORK

A large Japanese cross-sectional study ( $n = 30,000$ ) found that 28% slept less than 6 hours per night and 65% slept less than 7 hours per night despite 80% claiming adequate sleep [4]. Logistic regression identified short sleep duration ( $<6$  hours) as associated with being female, younger age, urban living, unemployment, poor health, lack of exercise, and irregular eating [4]. A review of Asian literature linked insufficient sleep to high stress, high BMI/obesity, and depression [5]. A study of Lithuanian university students ( $n = 405$ ) found poor sleep quality, especially among medical students, to be associated with high academic demands, anxiety, limited leisure time, and academic dissatisfaction [6]. These findings suggest that interventions targeting stress, leisure time, exercise, urban planning, mental health, and other lifestyle factors could potentially improve sleep [4], [6]. However, existing research lacks exploration of newer lifestyle factors, such as the impact of technology and social media on sleep habits. Future research is needed to investigate how emerging habits like nighttime electronic device use, pre-sleep social media

<sup>1</sup><https://github.com/johnkimtech/sleep>

engagement, and internet overuse might contribute to sleep disturbances.

### III. DATA ACQUISITION & PREPROCESSING

The DreamHarmony Sleep Survey was an integral part of our study, designed to explore the multifaceted relationship between lifestyle habits and sleep quality. To capture a diverse set of responses, the survey was made available in multiple languages including Korean, English, Vietnamese, and Bengali.

#### A. Data collection through online survey

The DreamHarmony Sleep Survey was meticulously designed to delve into the myriad factors affecting sleep quality and patterns. This comprehensive survey aimed to unravel the complex interplay between lifestyle habits and sleep, potentially leading to enhancements in sleep quality and overall life satisfaction.

*Survey Structure and Distribution:* The survey was structured into distinct sections: demographics, lifestyle, and sleep habits. It was distributed via Google Forms [7] through the authors' social networks, encompassing friends, colleagues, and family members. While this method facilitated rapid data collection, we acknowledge the potential for bias and sample imbalance due to the nature of the distribution channels.

*Demographics:* Participants were queried about their age, gender, education level, and occupation. This demographic information was crucial to understanding the contextual background of each respondent.

*Lifestyle and Sleep Habits:* The lifestyle section encompassed questions regarding exercise frequency, device usage, and screen time before sleep. In the sleep habits section, we focused on gathering detailed information about participants' sleep patterns. Key questions included:

- **Bedtime and Wake-up Time:** These questions helped calculate the actual duration of nighttime sleep, a vital metric for assessing sleep quality.
- **Sleep Quality:** Respondents rated their overall sleep quality, providing us with a subjective assessment of their sleep experience.
- **Sleep Onset:** We inquired about the time it typically takes for respondents to fall asleep, aiming to understand sleep onset difficulties.
- **Sleep Medication:** This question ascertained whether participants use any medication to aid their sleep, which is indicative of sleep quality issues.
- **Sleep Disturbances:** The frequency of disturbances such as waking up during the night or experiencing restless sleep was also captured.

*Contributions:* Our survey, though limited by its distribution method, provides valuable insights into sleep quality and its influencing factors. The subsequent statistical and machine learning analyses contribute to a nuanced understanding of sleep dynamics, laying the groundwork for future research in this field.

#### B. Preprocessing: Translation and Aggregation

#### IV. ANALYSIS

#### V. HYPOTHESIS TESTING

#### VI. MODELING

#### VII. CONCLUSION

#### VIII. FUTURE WORK

#### IX. RELATED WORK

#### ACKNOWLEDGMENT

#### REFERENCES

- [1] "Pearson correlation coefficient - wikipedia," [https://en.wikipedia.org/wiki/Pearson\\_correlation\\_coefficient](https://en.wikipedia.org/wiki/Pearson_correlation_coefficient), (Accessed on 12/15/2023).
- [2] "Spearman's rank correlation coefficient - wikipedia," [https://en.wikipedia.org/wiki/Spearman%27s\\_rank\\_correlation\\_coefficient](https://en.wikipedia.org/wiki/Spearman%27s_rank_correlation_coefficient), (Accessed on 12/15/2023).
- [3] "Chi-squared test - wikipedia," [https://en.wikipedia.org/wiki/Chi-squared\\_test](https://en.wikipedia.org/wiki/Chi-squared_test), (Accessed on 12/15/2023).
- [4] T. Ohida, A. Kamal, M. Uchiyama, K. Kim, S. Takemura, T. Sone, and T. Ishii, "The Influence of Lifestyle and Health Status Factors on Sleep Loss Among the Japanese General Population," *Sleep*, vol. 24, no. 3, pp. 333–338, 05 2001. [Online]. Available: <https://doi.org/10.1093/sleep/24.3.333>
- [5] B. Gómez-González, E. Domínguez-Salazar, G. Hurtado-Alvarado, E. Esqueda-Leon, R. Santana-Miranda, J. A. Rojas-Zamorano, and J. Velázquez-Moctezuma, "Role of sleep in the regulation of the immune system and the pituitary hormones," *Annals of the New York Academy of Sciences*, vol. 1261, no. 1, pp. 97–106, 2012. [Online]. Available: <https://nyaspubs.onlinelibrary.wiley.com/doi/abs/10.1111/j.1749-6632.2012.06616.x>
- [6] E. Preišėgolavičiūtė, D. Leskauskas, and V. Adomaitienė, "Associations of quality of sleep with lifestyle factors and profile of studies among lithuanian students," *Medicina (Kaunas, Lithuania)*, vol. 46, pp. 482–9, 07 2010.
- [7] "Dreamharmony sleep survey," [https://docs.google.com/forms/d/e/1FAIpQLSfJD8RAKNcRU6fjVxgMPsump2wiMIZPIH0QK3\\_wU6qwMxOEjw/viewform](https://docs.google.com/forms/d/e/1FAIpQLSfJD8RAKNcRU6fjVxgMPsump2wiMIZPIH0QK3_wU6qwMxOEjw/viewform), (Accessed on 12/15/2023).