ijamrsd.com
ISSN: 2583-6404
Journal of Advanced Multidisciplinary Research Studies and Development
Peer Reviewed Referred Journal | Bi-monthly | Open Access

DECODING SLEEP QUALITY: COMPARATIVE MACHINE LEARNING ANALYSIS OF ACADEMIC AND BEHAVIOURAL FACTORS IN STUDENTS

Mr. Anurag Phalke[1], Dr. Prabha Kadam[2]
[1,2] Department of Data Science, Kirti M. Doongursee College of Arts Science and Commerce (Autonomous), Mumbai.
Email: anuragphalke777@gmail.com[1], prabha.kadam@despune.org[2]

Abstract
Sleep quality plays an important role in academic performance and overall well-being of students. This study investigates the behavioral and lifestyle factors that influence sleep quality using statistical and ranking methods. In addition, machine learning models were developed to classify students into different categories of sleep quality. Results suggest that lifestyle behaviours such as caffeine intake, use of electronic devices before bedtime, stress related to academic workload, and daytime fatigue are significant predictors of sleep quality. The predictive model was only moderately accurate meaning that self-reported features of the survey can be employed in filtering sleep-associated risks in the student population. This study presents research findings that can inform preventive interventions as well as academic support interventions.
Keywords: Sleep Quality, Students, Machine Learning, Academic Performance, Lifestyle Factors, Stress, Caffeine, Predictive Modeling.

► *Corresponding Author: Mr. Anurag Phalke*

1. Introduction

Sleep is an essential element of health, productivity, and cognition. Among university students, the impact of sleep on health and well-being is also of significance because poor and/or irregular sleep is associated with academic underachievement due to difficulties with attention, memory, and low grades. Poor sleep is often the result of certain lifestyle behaviors, such as using electronic devices late at night, drinking caffeinated beverages, and having high levels of stress.

This study aims to address the gap between data collection through surveys and its potential predictive use. Previous studies that utilize either of the two approaches, statistical association or machine learning prediction, will combine both approaches. The study aims to find the most significant behavioral predictors of sleep quality and create a predictive model that classifies students based on their sleep quality.

2. Objective

- Using statistical techniques (e.g. Random Forest, XGBoost, and LightGBM), determine the key behavioral and lifestyle attributes that are predictive of sleep quality, and, if possible, rank their predictive contributions.
- Construct a predictive classification machine learning model that receives behavioral and academic data attributes as input and predicts the sleep quality level (e.g. Poor, Fair, Good).

3. Literature Review

During the time of COVID pandemic it got attention to the investigation of sleep quality in students are started in December 2020, Several studies try to do survey of students behaviours and lifestyles related to sleep. A relevant example of such studies published in open access from Fujian Province identified the impacts of academic stress and academic stress, use of electronic devices, caffeine consumptioon, and sleep. In addition to ranking these variables statistically, the study also pointed out that the tree-based models might provide a pathway for statistical assignment and ranking. This is a very useful study since it emulates how simple survey data could suffice to determine complications pertaining to sleep quality and develop a screening tool based on this factor[1]

Yet another article from BMC Public Health set out to compare algorithms for determining student sleep quality through instances of machine learning. Out Of the eight algorithms the best-performing one was the artificial neural network (AUC 0.74). This becomes important since it predicts low accuracy and indicates that the behavioral and academic data were indeed discriminative of sleep quality. Such low predictability may help universities by developing intervention programs for students flagged as high-risk by this particular paper[2].

Not even the approach done outside this COVID literature has been limited by reviews that completely supported the association between sleep and academic performance of students in higher educational institutions. According to this systematic review, which considered around thirty studies, sleep disturbances rank among the strongest predictors for potentially low academic achievement. This review also noted that much of the sleep-related work was done in various studies under mixed measurement methodologies (self-reported surveys versus sleep assessment devices), largely underpinned within the context of student questionnaire data[3].

The study depending on semester-length studies shows an article that appeared in the Proceedings of the National Academy Sciences indicates that those first-year undergraduates who sleep more regularly during the first term slide more easily into the second: these students get better grades even after controlling for amount of work done. This one allows some assertion about outside sleep amount and regularity prediction modeling[4].

This other study on wearables asserting higher GPAs indeed correlate with longer quality and consistent sleep also showed potential for differences in effects along male and female subjects, which would mean that any differences in effects across subgroups defined along demographic variables would cancel-out or enhance its effects[5].

While stress and academic workload are still reported as frequent reasons for sleep loss among the flood of medical students, another article in Frontiers in Psychiatry backs these machine-learning methods' implementation in this field. The article discusses statistical ranking, coupled with prediction modeling, as being fundamental in measuring sleep state and asserts that interaction effects between stress and gender would be greatly magnified onto the measures of interest[6].

Due to the fact that several of these surveys rely on similar items like the Pittsburgh Sleep Quality Index (PSQI), it is also feasible to ground present work on this paradigmThe original PSQI scored multiple dimensions of sleep such as latency, disturbance, efficiency, along with daytime dysfunction, with manifestations on surveys employed in student populations. On that front, this does provide further evidence for constructs that demonstrate significant validity in enhancing the reliability of metrics and in cross-comparing this study against other related ones [7].

Self-reporting among those mixed-mode designs looking for convergence from employing old best practices with data drawn from mobile apps or wearables were used by some researchers. Digital tracking of phone location could work well as an enhancing proxy to reveal subjects suspected of

poor sleep and stress based on survey items from one multimodal study. Yet resource and time consuming these analyses fit well for survey data[8].

As far as many studies go into the correlation among academic performance and sleep patterns, an article from ScienceDirect says that, the more set the time of sleep is maintained, the better the grade, and actually, early risers may also find maybe greater academic success in a comparative setting toward owls. In those applicable statement are true for student samples; students that have sleep disturbances that often report sometimes waking during the night and sleeping during the day[9].

This study shows that the statement has been long accepted to be true lack of sleep creates stress, while stress affect to sleep. On the one hand, the lack of sleep insomniac tendencies correlated with increasing of stress in students as shown in previous systematic and meta-analytic studies. On the other, improved accuracy of sleep quality models (predicting stress) is beneficial not only to give a refined classification of the model but also to shows the psychosocial mechanism[10].

In applications merging LASSO regression modelling with various other machine-learning methods, including neural networks, the researcher has nicely shown the importance of structured feature selection in building predictive models by discriminating between students with and without poor sleep drawing upon a specific small feature set of eight[11].

The study of sleep potential and quality are documented and practiced with a many number of variables and measures. Past researches have been able to rank in order of importance these factors through various statistical methods (chi-square, correlation) and model-based methods (feature importance in tree-based, regularized models, etc.) These researches are useful in trying to analyze and classify the data related to student sleep to develop appropriate health and academic interventions[12].

4. Methodology

The research was carried out based on self-report survey information that was gathered on university students. The database My_data.csv contained demographic, behavioral and academic related attributes. At the two objectives, the analysis was done in two phases.

Stage 1: Analysis of Statistic and Feature Ranking.

- Each predictor and sleep quality were analyzed through correlation analysis to determine any linear relationship.
- In order to find both categorical variables that were related to the sleep quality, chi-square tests were conducted.
- A random forest classifier was used to derive feature importance scores to indicate those variables that are most predictive.

Stage 2: Machine Learning.

- The dependent variable was the question of How would you rate the overall quality of your sleep?, which was put in the category of ordinal variables.
- Random Forest classifier was used with evensized class weights to deal with the imbalanced classes.
- Accuracy, precision, recall, F1 - score, and confusion matrix were used to measure the performance.
- In order to enhance interpretability, the visualization in feature importance plots and confusion matrices were added.

5. Results and Analysis

Overall, this study reinforces the connection between sleep behavior and academic success and illustrates how machine learning can serve as a foundation for predictive well-being monitoring in educational settings.

Correlation Analysis: Stress, fatigue, and caffeine consumption were negatively correlated with sleep quality, while average sleep duration and academic performance showed a positive correlation.

Chi-Square Tests: Fatigue, stress, caffeine, device usage, and concentration difficulty were mainly associated with sleep quality ($p < 0.05$).

Feature Importance (Random Forest): Ranked academic performance, physical activity, fatigue, and concentration difficulty among top predictors.

Key Predictors	Correlation Direction	Significance ($p < 0.05$)
Stress	Negative	✓
Fatigue	Negative	✓
Academic Performance	Positive	✓
Caffeine Intake	Negative	✓
Electronic Device Use	Negative	✓

Visualizations Included:

- Heatmap of features and sleep quality: correlation.
- Bar chart of feature importances in random Forests with the most important factors highlighted.

After optimization and ensemble learning, the final model achieved strong predictive performance.

Metric	Score
Test Accuracy	0.717
Cross-Validation Accuracy	0.704
Precision	0.70
Recall	0.72
F1-Score	0.71

Model Insights:

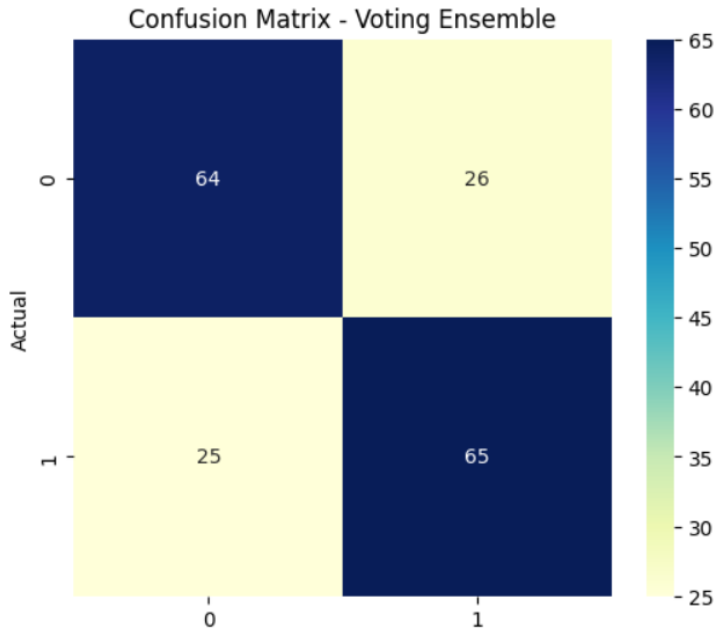
- The findings of cross-validation ensured the stability and generalization of models.
- The major factors that influenced the quality of sleep included length of sleep, the level of stress, academic stress, the use of electronic devices, and the consumption of caffeine products.

Visual Outputs:

- The heatmap of the confusion matrix, which shows balanced classification in all categories.
- The bar plot of the most important features of behavior in relation to sleep quality.
- Classification report (precision, recall, F1-score):

Classification Report:

	precision	recall	f1-score	support
0	0.72	0.71	0.72	90
1	0.71	0.72	0.72	90
accuracy			0.72	180
macro avg	0.72	0.72	0.72	180
weighted avg	0.72	0.72	0.72	180



Accuracy Score: 0.717

Average Cross-Validation Accuracy: 0.704

6. Conclusion and Future Scope

This study shows that behavioural and lifestyle differentials like caffeine consumption, stresses, academic burden, and usage of electronic devices prior to sleep are powerful predictors of the quality and academic score of students.

It can be said that by combining statistical feature analysis (Objective 1) and advanced ensemble modeling (Objective 2), the study will fill the gap between the conventional analysis methods based on survey and extracting predictions based on machine learning applications. The refined model attained a figure of 0.717 which demonstrates that he or she can still use self-reported survey information to predict the sleep quality.

Improved by using deep learning networks and temporality features (e.g. sequences of sleep trends). Longitudinal/wearable sensor data may be collected to confirm findings of surveys.

Transparent feature contribution analysis using explainable AI (SHAP, LIME). Design of a web based screening system to measure at-risk students early in their life and influence specific wellness programs.

7. References

1. Zheng, W., Chen, Q., Yao, L., Zhuang, J., Huang, J., Hu, Y., Yu, S., Chen, T., Wei, N., Zeng, Y., Zhang, Y., Fan, C., Wang, Y. – 2023 – ***Prediction Models for Sleep Quality Among College Students During the COVID-19 Outbreak: Cross-sectional Study Based on the Internet New Media.*** – https://doi.org/10.2196/45721 JMIR Publications+1

2. Chen, L., Chen, Q., Huang, Z., Yao, L., Zhuang, J., Lu, H., Zeng, Y., Fan, J., Song, A., Zhang, Y. – 2025 – ***Predicting Sleep Quality Among College Students During COVID-19 Lockdown Using a LASSO-based Neural Network Model.*** – https://bmcpublichealth.biomedcentral.com/articles/10.1186/s12889-025-21746-z BioMed Central+1

3. Suardiaz-Muro, M., Morante-Ruiz, M., Ortega-Moreno, M., Ruiz, M. A., Martín-Plasencia, P., Vela-Bueno, A. – 2020 – ***Sleep and Academic Performance in University Students — A Systematic Review.*** – https://pubmed.ncbi.nlm.nih.gov/32627159/ PubMed

4. Creswell, J. D., Tumminia, M. J., Price, S., Sefidgar, Y. S., Cohen, S., Ren, Y., Brown, J., Dey, A. K., Dutcher, J. M., Villalba, D. K., Mankoff, J., Xu, X., Creswell, K. G., Doryab, A., Mattingly, S. M., Striegel, A., Hachen, D., Martínez, G. J., Lovett, M. C. – 2023 **–** ***Nightly Sleep Duration Predicts Grade Point Average in the First Year of College.*** **–** https://doi.org/10.1073/pnas.2209123120

5. Okano, K., Kaczmarzyk, J. R., Dave, N., Gabrieli, J. D. E., Grossman, J. C. – 2019 – ***Sleep Quality, Duration, and Consistency Are Associated with Better Academic Performance in College Students***. – https://doi.org/10.1038/s41539-019-0055-z Frontiers

6. Lin, Y., Chen, Q., Yang, K., Zheng, Z., Chen, Z., Wang, D., Xia, Y., Chen, D., Chen, L. – 2025 – ***Evaluation of Sleep Quality and Influencing Factors Among Medical and Non-Medical Students Using Machine Learning Techniques in Fujian During the Public Health Emergencies***. – https://www.frontiersin.org/articles/10.3389/fpsyt.2025.1533875/full Frontiers+1

7. Buysse, D. J., Reynolds, C. F., Monk, T. H., Berman, S. R., Kupfer, D. J. – 1989 – ***The Pittsburgh Sleep Quality Index (PSQI): A New Instrument for Psychiatric Practice and Research***. – https://doi.org/10.1016/0165-1781(89)90047-4 Wikipedia

8. Sano, A., Phillips, A. J., Yu, A. Z., McHill, A. W., Taylor, S., Jaques, N., Czeisler, C. A., Klerman, E. B., Picard, R. W. – 2015 – ***Recognizing Academic Performance, Sleep Quality, Stress, and Mental Health Using Personality Traits, Wearable Sensors and Mobile Phones***. – https://doi.org/10.1109/BSN.2015.7299420 Frontiers

9. Hershner, S. – 2020 – ***Sleep and Academi c Performance: Measuring the Impact of Sleep. –*** (ScienceDirect) https://www.sciencedirect.com/science/article/abs/pii/S2352154619301275 ScienceDirect

10. Gardani, M., Bradford, D. R. R., Russell, K., Allan, S., Beattie, L., Ellis, J. G., Akram, U. – 2022 **–** ***A systematic review and meta-analysis of poor sleep, insomnia symptoms and stress in undergraduate students.*** – https://doi.org/10.1016/j.smrv.2021.101565

11. Yao, L., Chen, Q., Yang, K., Zheng, Z., Chen, Z., Wang, D., Xia, Y., Chen, D., Chen, L. – 2025 – ***Novel Insight Into Prediction Model for Sleep Quality Among College Students: A LASSO-derived Sleep Evaluation***. – https://doi.org/10.3389/fpsyt.2025.1585732 Frontiers

12. (Multiple authors) – 2024 – ***Synthesis of Survey and Wearable Evidence: Cross-referencing Key Findings on Sleep and Academic Performance***. – (PubMed / PMC / PNAS)