



Comparison of Support Vector Machine (SVM) and K-Nearest Neighbors (KNN) Methods in Classifying University Students' Sleep Quality Based on Academic Assignment Workload

Mukhammad Rasya Azaka Makruf¹, Muhammad Iqbal Hafidz², Oktasae Lawusya Emka³,
Alfina Muna Maitsa⁴, Najwa Lailatun Nisa⁵, Mohammad Fadhel Putra Naue⁶

^{1,2,3}Department of Data Science, UPN "Veteran" Jawa Timur, Indonesia

⁴Department of Food Industry Quality Assurance, Politeknik AKA Bogor, Indonesia

⁵Department of Data Science Technology, Universitas Airlangga, Indonesia

⁶Department of Islamic Banking, Institut Agama Islam Manado, Indonesia

¹25083010073@student.upnjatim.ac.id, ²25083010047@student.upnjatim.ac.id, ³25083010065@student.upnjatim.ac.id,
⁴alfinamunamaita@gmail.com, ⁵najwa.lailatun.nisa-2025@ftmm.unair.ac.id, ⁶moh.25142015@iain-manado.ac.id,

ABSTRACT

A heavy academic assignment workload in higher education can shorten students' nightly sleep and contribute to the accumulation of sleep debt. This study aims to analyze the association between academic assignment workload and rest patterns, and to compare the performance of the Support Vector Machine (SVM) and K-Nearest Neighbors (KNN) algorithms in classifying university students' sleep quality. Primary data were collected through a structured questionnaire; of the 159 responses received, 150 respondents were selected according to per-faculty quotas and were distributed across 8 faculties. The models used 9 numeric attributes covering assignment workload and rest duration, with a binary target representing perceived sleep quality: Class 0 (Restful, $n=86$ or 57.3%) and Class 1 (Poor, $n=64$ or 42.7%). Correlation analysis showed a moderate negative association ($r = -0.44$) between daily assignment work duration and sleep hours on active lecture days. The primary evaluation used Nested Stratified 5-Fold Cross-Validation to prevent data leakage during hyperparameter search. The mean accuracy of SVM was 58.67% ($\pm 10.43\%$) and that of KNN was 51.33% ($\pm 7.67\%$), compared with 57.33% ($\pm 1.49\%$) for a majority-class baseline model. The SVM accuracy was only 1.33 percentage points above the baseline, a difference smaller than the between-fold standard deviation, whereas the KNN accuracy was below the baseline. In the supporting evaluation on a holdout test set ($n=30$), SVM achieved an accuracy of 66.67% (Macro F1-Score 64.11%; ROC-AUC 0.7647) and KNN 56.67% (Macro F1-Score 56.62%; ROC-AUC 0.5724), while the baseline reached 56.67%. Given the small sample size and overlapping confidence intervals, the difference between SVM and KNN is reported descriptively and is not interpreted as general superiority. The limited performance of both models is possibly related to the variability of subjective perceptions in questionnaire data, the small sample size, and overlapping characteristics between classes.

Keywords: Sleep Quality, Academic Assignment Workload, Support Vector Machine, K-Nearest Neighbors, Nested Cross-Validation, Baseline Classifier.

* *Penulis korespondensi.*

Alamat e-mail: 25083010073@student.upnjatim.ac.id
<http://doi.org/10.33005/jasid.v2i2.81>

I. Introduction

Sleep is an essential biological requirement for human life because it plays a role in memory consolidation, the restoration of neurocognitive function, and emotional regulation [1]. Chronic sleep deprivation is associated with reduced critical thinking capacity, slower information processing, emotional instability, and a higher risk of physical and mental health problems [1]. In modern higher education, nightly rest is often displaced by academic productivity demands, which makes the quality and duration of students' sleep a relevant issue to study.

Specifically, university students routinely face accumulating coursework, tight deadlines, and a heavy load of semester credit units (SKS). These conditions can push students to cut their daily sleep to complete academic obligations and trigger sleep debt [2], namely sleep duration on active lecture days that is cut short and then compensated for on weekends. Therefore, the relationship between academic assignment workload and sleep quality needs to be mapped more objectively, rather than only through qualitative assessment or simple linear assumptions. Machine learning, particularly classification algorithms such as Support Vector Machine (SVM) and K-Nearest Neighbors (KNN), is a computational approach that can be used to examine non-linear relationships between assignment workload and sleep quality.

The problem addressed here is that sleep quality in this study is measured through respondents' subjective assessments, while the predictors are questionnaire answers about assignment workload and sleep habits. Such data are self-reported, show high perceptual variation between individuals, and do not measure physiological conditions directly, so sleep quality classes may overlap in the feature space. With small questionnaire datasets, evaluation based on a single train-test split also yields unstable performance estimates, and accuracy values are difficult to interpret without a simple comparison such as a model that always predicts the majority class. Consequently, it remains unclear to what extent academic assignment workload data can serve as a predictor of students' sleep quality, and whether SVM or KNN performs better than the baseline.

Previous studies have used classification algorithms such as SVM and KNN to detect stress levels and rest quality, and most of them rely on objective physiological data obtained from wearable sensors or electrocardiogram signals [3], [4]. Sensor-based approaches can achieve high classification accuracy, but their application to large student populations is relatively limited because they require expensive instruments and complex technical involvement. In studies that use questionnaire-based self-report data, some still rely on a single train-test split without a cross-validation framework that is free from data leakage [5], and models are often not compared with a simple reference model (Majority Class Classifier), so claims of an algorithm's superiority remain insufficiently tested [5], [6]. Based on this, the present study aims to analyze the association between academic assignment workload and students' sleep patterns and to compare the performance of SVM and KNN in classifying sleep quality, using Nested Stratified 5-Fold Cross-Validation as the primary evaluation, a majority-class baseline as the comparator, and holdout testing as supporting evidence.

II. RESEARCH METHODOLOGY

The systematic stages of this study are represented through a workflow consisting of four main phases. These four phases are designed sequentially to build an optimal predictive model. The workflow begins with data collection and exploratory data analysis (EDA), continues with data preprocessing and standardization and the classification modeling process using machine learning algorithms (Support Vector Machine and K-Nearest Neighbors), and ends with the evaluation of model performance along with the visualization of its results.



A. Data Collection

The primary data in this study were obtained by distributing an online questionnaire to active students. Of the 159 responses collected (all of which indicated willingness to participate, and no duplicate student ID numbers were found), 150 respondents were selected through sampling based on per-faculty quotas with a fixed random seed (random_state=42): 45 respondents from the Faculty of Computer Science and 15 respondents from each of the seven other faculties. This composition follows quotas set by the researchers and does not represent the proportions of the student population. The dataset consists of 150 respondent records with a total of 15 parameters on academic assignment work behavior and rest patterns. Table 1 below presents part of the metadata of the main attributes used in this study.

ID	Parameters	Description	Data Type
1	Number of Academic Assignments	Average number of academic assignments that must be completed in one week.	Numeric
2	Deadline Range	Average time span (in days) given from the moment an assignment is issued until the submission deadline.	Numeric
3	Assignment Work Duration	Average time (in hours) spent per day on academic assignments.	Numeric
4	Sleep Hours (Weekdays)	Average nightly sleep duration on active lecture days (Monday–Friday).	Numeric
5	Frequency of Staying Up Late	Intensity (frequency) with which students fight off sleepiness or stay up all night to work on assignments.	Numeric
6	Number of Credits (SKS)	Estimated total semester credit units (SKS) taken by the respondent this semester.	Numeric
7	Class Tardiness	Frequency of arriving late to morning classes due to fatigue in the past week.	Numeric
8	Sleep Hours (Days Off)	Average nightly sleep duration on non-lecture days (Saturday–Sunday).	Numeric
9	Recovery Time	Estimated time (in days) the body needs to restore its sleep cycle and fitness after staying up late.	Numeric
10	Sleep Quality (Target)	Respondent's subjective assessment of their nighttime sleep quality after working on assignments.	Categorical
11	Assignment Working Time	Time range (hours/session) most often used by respondents to complete assignments.	Categorical
12	Effects of Sleep Deprivation	Most dominant physical or psychological effect felt on campus due to sleep deprivation.	Categorical

13	Assignment Pile-up Factor	Respondents' view on the main factor causing academic assignments to often pile up.	Categorical
14	Workload vs. Rest Ratio	Assessment of the ratio (comparison) between academic assignment volume and the availability of rest time.	Categorical
15	Drowsiness Coping Solution	Strategy or step most often taken by respondents to cope with drowsiness on campus.	Categorical

Table 1. Research Variables

B. Data Preprocessing

The predictor feature matrix was formed from 9 numeric attributes representing academic assignment workload and daily sleep patterns (Questions 1 to 9 of the questionnaire). Variable X1 is the average number of academic assignments per week. Variable X2 is the average time span of assignment deadlines in days. Variable X3 represents the time spent specifically on assignments outside class hours per day, in hours. Variable X4 is the average nightly sleep duration on active lecture days (Monday to Friday), in hours. Variable X5 records the frequency of being forced to sleep after midnight because of assignments in the past week. Variable X6 is the total semester credit units (SKS) taken in the current semester. Variable X7 records the frequency of arriving late to morning classes due to fatigue in a week. Variable X8 is the average nightly sleep duration on days off (Saturday and Sunday), in hours. Variable X9 represents the estimated recovery time needed by the body, in days, for physical condition to return to normal after staying up late.

The target variable (Y) was determined from respondents' subjective assessment of their nighttime sleep quality after completing piles of academic assignments (Question 10). The qualitative responses, consisting of four ordinal scales, were grouped into two binary classes to facilitate the classification task. The categories "Very Restful" and "Fairly Restful" were mapped to Class 0 (Restful sleep quality), comprising 86 respondents (57.3%). The categories "Not Very Restful" and "Very Poor" were mapped to Class 1 (Poor sleep quality), comprising 64 respondents (42.7%). To preserve modeling validity and avoid information leakage (target proxies), five additional qualitative questions on effects and coping strategies for drowsiness (Questions 11 to 15) were excluded from the predictor matrix.

Before training, all numeric attributes X1 ... X9 underwent missing-value imputation using the median, followed by scale adjustment using StandardScaler. No missing values were found in this dataset, but imputation was still included in the pipeline as a safeguard. Extreme values (outliers) were neither replaced nor removed because they could not be confirmed as input errors; one answer of 22 hours per day for assignment work duration was retained and recorded as a limitation. The scale adjustment formula is shown in Equation (1):

$$z = \frac{x - \mu}{\sigma} \quad (1)$$

In Equation (1), z represents the standardized feature value, x is the raw value, μ is the feature mean in the training data, and σ is the feature standard deviation. This entire data transformation process was strictly isolated within the validation pipeline to ensure that information from the test data never leaked into the training process.

C. Classification Modeling

The classification stage aims to predict students' sleep quality labels based on academic workload parameters using two machine learning approaches.

1. K-Nearest Neighbors (KNN)

The KNN algorithm classifies data based on the spatial proximity between observation points. A new data point is predicted from the majority label of its k nearest neighbors (the value of k was selected from {1, 3, 5, 7, 9, 13, 17, 21, 25} through GridSearchCV). The metric used to measure the distance between respondents is the Euclidean distance, which calculates the direct

straight-line distance in n-dimensional space [5]. Its formulation is presented in Equation (2).

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (2)$$

Where:

$d(x, y)$: Euclidean distance between data points x and y

n : Number of dimensions (observation features)

x_i, y_i : Value of the i -th attribute of data x and y

2. Support Vector Machine (SVM)

The SVM model aims to find the optimal decision boundary (hyperplane) that maximizes the margin separating the sleep quality classes. The kernels evaluated were the linear kernel and the Radial Basis Function (RBF) kernel; the RBF kernel projects the data into a higher-dimensional space so that it can form non-linear decision boundaries on overlapping data. Hyperparameters were selected through GridSearchCV with a search space of the linear kernel ($C = 0.1, 1, 10, 100$) and the RBF kernel ($C = 1, 10, 100$ with $\gamma = \text{'scale'}$ or 0.1). Mathematically, the RBF kernel transformation is formulated in Equation (3).

$$K(x, x') = \exp(-\gamma \|x - x'\|^2) \quad (3)$$

Where:

$K(x, x')$: RBF kernel function

$\|x - x'\|^2$: Squared Euclidean distance between two feature vectors

γ : Hyperparameter that controls the range of influence of a single training data point

D. Evaluation

The primary evaluation was carried out with Nested Stratified 5-Fold Cross-Validation. The outer loop (stratified 5-fold, seed 123) was used to estimate performance, while the inner loop (stratified 5-fold, seed 42) was used to select hyperparameters using accuracy as the metric, so that outer-fold data were not used in hyperparameter selection. Imputation, standardization, and hyperparameter search were run inside the pipeline only on the training data of each fold. As a comparator, a baseline in the form of a DummyClassifier with the most_frequent strategy (always predicting the majority class) was trained only on the training data of the outer fold. As supporting evidence, a stratified 80:20 holdout split (random_state=42) was performed, yielding 120 training samples and 30 test samples, with hyperparameters selected through GridSearchCV on the training data. Results are reported as the mean and standard deviation across folds; 95% Wilson confidence intervals were computed as a descriptive summary and not as a significance test.

Classification effectiveness was measured using the Confusion Matrix, which yields evaluation metrics such as Accuracy, Precision, Recall, and F1-Score. The matrix components consist of True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN).

Accuracy measures the percentage of correct predictions out of all test data. Its formulation is presented in Equation (4).

$$Q = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

Precision measures the ratio of correct positive predictions to all data predicted as positive, while Recall measures the model's ability to find all actual positive data. F1-Score is used as the

harmonic mean of Precision and Recall, which is particularly useful for data with class imbalance. The formulations of these three metrics are presented in Equations (5) and (6).

$$Precision = \frac{TP}{TP+FP}, Recall = \frac{TP}{TP+FN} \quad (5)$$

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (6)$$

Macro F1-Score is the average of the F1-Scores of both classes, while Balanced Accuracy is the average of the Recall of both classes. ROC-AUC was computed on the holdout test data from the decision function scores (SVM) and class probabilities (KNN); because the baseline predicts only one class, the baseline ROC-AUC is undefined.

III. RESULTS AND DISCUSSION

A. Descriptive Characteristics of Respondents

Descriptive statistical analysis of the 9 numeric variables across 150 respondents provides a quantitative overview of students' assignment workload and sleep patterns. The mean (μ), standard deviation (σ), minimum, median, and maximum values are summarized in Table 2.

Predictor Attribute	Mean (μ)	Std. Deviation (σ)	Min	Median	Max
X1: Number of Academic Assignments (assignments/week)	4.41	1.78	1.00	4.00	10.00
X2: Deadline Range (days)	6.03	1.64	1.00	7.00	7.00
X3: Assignment Work Hours (hours/day)	4.33	2.68	1.00	4.00	22.00
X4: Sleep Hours on Lecture Days (hours/night)	5.57	1.84	1.00	5.50	12.00
X5: Frequency of Staying Up Late (times/week)	2.73	2.06	0.00	3.00	7.00
X6: Number of Credits (SKS) This Semester	20.99	3.82	3.00	22.00	24.00
X7: Class Tardiness (times/week)	0.34	0.85	0.00	0.00	4.00
X8: Sleep Hours on Days Off (hours/night)	7.35	2.34	1.00	7.50	15.00
X9: Fitness Recovery Time (days)	1.51	1.58	0.00	1.00	8.00

Table 2. Respondent Characteristics

The data in Table 2 show that students receive on average 4.41 academic assignments per week and spend 4.33 hours per day on assignments outside lecture hours. The mean nightly sleep duration on active lecture days is 5.57 hours, shorter than the 7 to 9 hours generally recommended for young adults. On weekend days off, mean sleep duration rises to 7.35 hours, a difference of

about 1.78 hours. This pattern is consistent with the possibility of sleep debt recovery on weekends, but the difference is descriptive and has not been tested statistically. The high maximum values on several variables, for example an assignment work duration of 22 hours per day, were retained as they are because they could not be confirmed as input errors.

B. Pearson Correlation Analysis among Variables

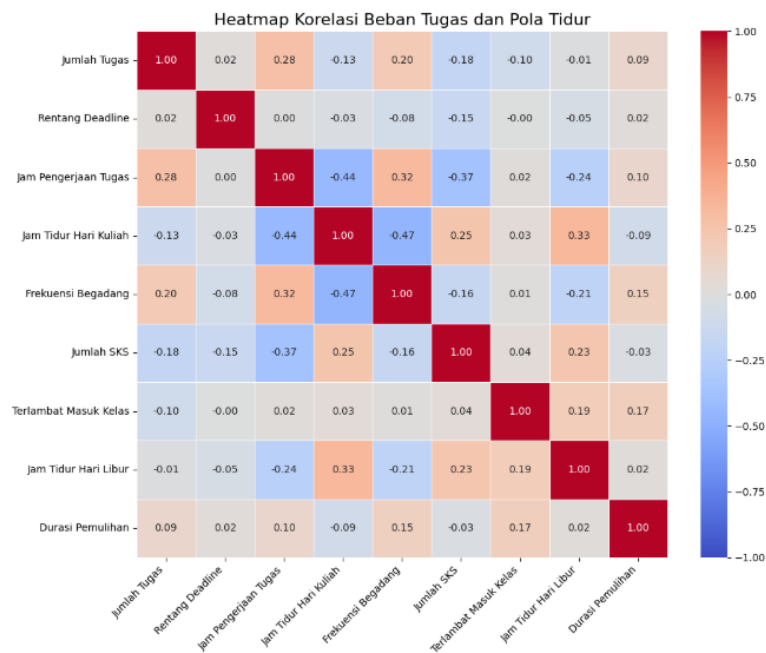


Figure 1. Correlation Heatmap

The correlation matrix analysis reveals several patterns of relationships among variables. Between daily assignment work duration (X3) and sleep hours on active lecture days (X4), a moderate negative correlation of $r = -0.44$ was found. This value indicates a tendency for respondents who report longer assignment work time to report shorter sleep on lecture days; this association cannot be interpreted as a causal relationship.

In addition, assignment work duration (X3) is positively associated with the weekly frequency of staying up late (X5) at $r = +0.32$, and with the number of academic assignments (X1) at $r = +0.28$. Sleep hours on active lecture days (X4) are also positively correlated, at $r = +0.33$, with sleep hours on days off (X8). Of the 36 variable pairs, 31 have $|r| < 0.30$ and the largest absolute correlation is 0.47, so correlations among variables are generally weak to moderate and there is no indication of strong multicollinearity. All nine variables were retained in the modeling. These generally weak to moderate correlations are also relevant context for the classification results in the following sections.

C. Primary Model Performance Evaluation (Nested Stratified 5-Fold Cross-Validation)

The comparative performance test of SVM, KNN, and the Baseline Classifier on the entire dataset ($n=150$) using Nested Stratified 5-Fold Cross-Validation produced the performance summary in Table 3. The majority baseline in each outer fold was determined from the training data of that fold.

Performance Evaluation Metric	Baseline (Majority)	K-Nearest Neighbors (KNN)	Support Vector Machine (SVM)
-------------------------------	---------------------	---------------------------	------------------------------

Mean Accuracy (Mean $\pm$ SD)	57.33% $\pm$ 1.49%	51.33% $\pm$ 7.67%	58.67% $\pm$ 10.43%
Macro F1-Score (Mean $\pm$ SD)	36.44% $\pm$ 0.59%	47.54% $\pm$ 11.41%	55.69% $\pm$ 10.43%
Balanced Accuracy (Mean $\pm$ SD)	50.00% $\pm$ 0.00%	49.29% $\pm$ 8.74%	56.35% $\pm$ 10.24%
Total Correct Predictions (OOF)	86 / 150	77 / 150	88 / 150
Out-of-Fold (OOF) Accuracy	57.33%	51.33%	58.67%
Wilson 95% CI (OOF Accuracy)	[49.33% – 64.97%]	[43.40% – 59.20%]	[50.67% – 66.23%]

Table 3. Nested Cross-Validation Results

The Nested CV results in Table 3 constitute the primary evidence on the predictive capacity of both algorithms. The mean accuracy of SVM is 58.67% with a standard deviation of 10.43%, or 1.33 percentage points above the majority baseline (57.33%). This difference is much smaller than the between-fold standard deviation, and at the per-fold level the SVM accuracy is higher than the baseline in 3 of 5 folds and lower in 2 folds. On metrics that account for both classes, SVM obtains a Macro F1-Score of 55.69% and a Balanced Accuracy of 56.35%, compared with 36.44% and 50.00% for the baseline. The baseline Macro F1-Score is mechanically low because the baseline predicts only one class, so Balanced Accuracy is the more appropriate comparator; the difference between the SVM Balanced Accuracy and the baseline (6.35 percentage points) still lies within the between-fold standard deviation (10.24%). These results indicate that SVM may capture some patterns in the data, but the evidence is limited and not sufficient to conclude that the model learns meaningful patterns consistently.

Meanwhile, KNN records a mean accuracy of 51.33% with a standard deviation of 7.67%, which is 6.00 percentage points below the baseline, and it is lower than the baseline in 4 of 5 folds. The KNN Balanced Accuracy of 49.29% is around the level of random guessing (50%). The selected value of k varied across folds ($k = 1, 5, 13, 13$, and 1), as did the SVM choices (RBF kernel in four folds and linear kernel in one fold), which indicates that hyperparameter selection is unstable on this sample. The low KNN performance is possibly related to a noisy feature space and class overlap in questionnaire data [5], which makes label assignment based on nearest neighbors prone to misclassification; this conjecture has not been tested directly.

The difference in mean accuracy between SVM and KNN in Nested CV is 7.33 percentage points, smaller than the largest between-fold standard deviation ($SD = 10.43\%$). At the per-fold level, SVM is higher than KNN in 3 folds and equal in 2 folds, and is never lower. Because there are only five folds that are not fully independent and the models share part of the training data, this difference is interpreted descriptively on the observed sample, neither as a significance test nor as evidence of general superiority of SVM [7], [10].

D. Supporting Evaluation (80:20 Holdout Test)

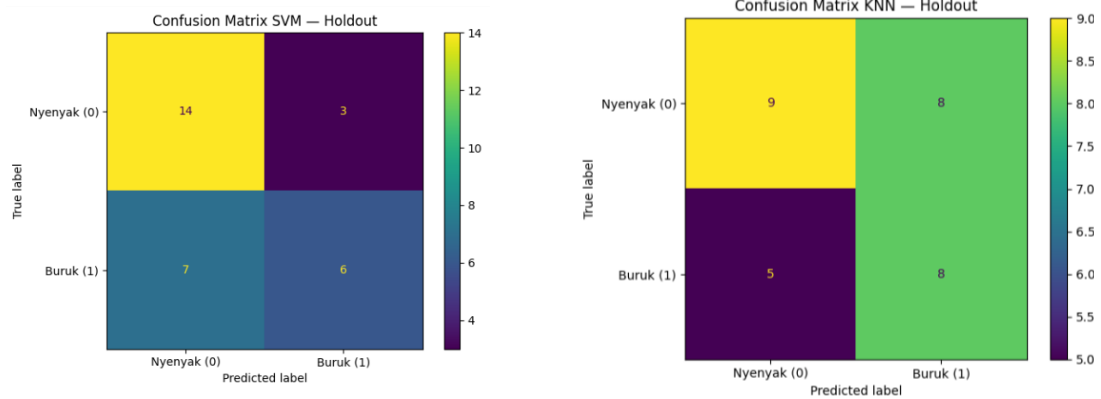


Figure 2. Confusion Matrices of SVM and KNN

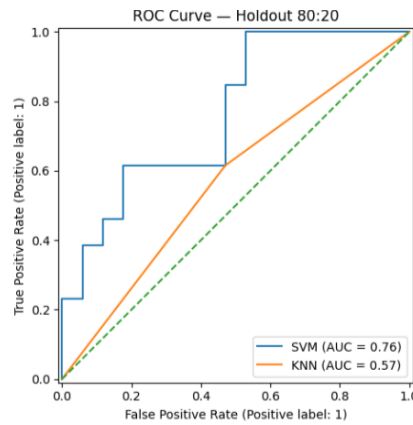


Figure 3. ROC Curve

In the independent holdout test using 30 test samples (17 Class 0 samples and 13 Class 1 samples), the best hyperparameters selected through GridSearchCV on the training data ($n=120$) were SVM ($C = 1$, $\gamma = 0.1$, RBF) and KNN ($k = 1$). The evaluation metrics of the independent holdout test are summarized in Table 4.

Performance Evaluation Metric	Baseline (Majority)	K-Nearest Neighbors (KNN)	Support Vector Machine (SVM)
Number of Correct Predictions	17 / 30	17 / 30	20 / 30
Test Accuracy	56.67%	56.67%	66.67%
Macro F1-Score	36.17%	56.62%	64.11%
Balanced Accuracy	50.00%	57.24%	64.25%
Area Under ROC (ROC-AUC)	Undefined	0.5724	0.7647
Wilson 95% CI (Test Accuracy)	[39.20% – 72.62%]	[39.20% – 72.62%]	[48.78% – 80.77%]

Table 4. Summary of Performance Evaluation Metrics

The classification report on the test data presents class-specific performance. For the SVM model, of the 17 actual Class 0 (Restful) samples, 14 were correctly predicted ($TN=14$) and 3 were incorrectly predicted ($FP=3$). Of the 13 actual Class 1 (Poor) samples, 6 were correctly

predicted (TP=6) and 7 were incorrectly predicted (FN=7). These results give a Class 0 Precision of 66.67%, a Class 0 Recall of 82.35%, and a Class 0 F1-Score of 73.68%. For Class 1, SVM obtains a Precision of 66.67%, a Recall of 46.15%, and an F1-Score of 54.55%. A Class 1 Recall of 46.15% means that more than half of the respondents with poor sleep quality were not detected by SVM. The SVM ROC-AUC of 0.7647 indicates better ranking ability than KNN on this test set, but it was computed from only 30 samples and is therefore tentative.

For the KNN model ($k=1$), of the 17 Class 0 samples, 9 were correctly predicted (TN=9) and 8 were incorrectly predicted (FP=8). Of the 13 Class 1 samples, 8 were correctly predicted (TP=8) and 5 were incorrectly predicted (FN=5). KNN yields a Class 0 Precision of 64.29%, a Class 0 Recall of 52.94%, and a Class 0 F1-Score of 58.06%. For Class 1, KNN obtains a Precision of 50.00%, a Recall of 61.54%, and an F1-Score of 55.17%, with a ROC-AUC of 0.5724.

With a test set of $n=30$, each prediction is equivalent to 3.33 percentage points of accuracy. The difference of 3 correct predictions between SVM (20 correct) and KNN (17 correct) is equivalent to 10.00 percentage points. The 95% Wilson confidence intervals of SVM ([48.78% – 80.77%]) and KNN ([39.20% – 72.62%]) overlap widely, and the SVM interval also covers the baseline accuracy (56.67%). KNN obtains the same number of correct predictions as the baseline (17/30). The SVM holdout accuracy (66.67%) is higher than the mean Nested CV accuracy (58.67%); this difference possibly reflects variation in the data split on a small test sample. The holdout results are therefore treated as supporting evidence whose direction is consistent with Nested CV (SVM higher than KNN), not as evidence of general superiority [7].

E. Comparative Discussion and Scientific Implications

The findings of this study offer a picture of applying machine learning to subjective questionnaire data. Nested CV accuracy ranges from about 51% to 59%, close to the majority baseline accuracy (57.33%), while the highest holdout accuracy of 66.67% was obtained on only 30 test samples. Accordingly, the evidence that both models learn meaningful patterns from the questionnaire features remains limited. The following factors are possible explanations that have not been tested, not causes that have been proven:

The first factor is the possibly high within-class variability (high intra-class variance) [8], [9]. The assessment of sleep quality through a questionnaire is perceptual. Two students with the same objective academic attributes, for example both taking 22 SKS, working on assignments 5 hours per day, and sleeping 5 hours per night, may give opposite subjective assessments. A student with high physical resilience may rate 5 hours of sleep as “Fairly Restful”, while another student who experiences chronic fatigue categorizes it as “Very Poor”. Psychological mediating variables such as anxiety level, sleep efficiency, and stress coping mechanisms are not captured in the standard academic questionnaire attributes [8].

The second factor is the limited sample size ($n=150$) relative to the dimensionality of the feature space [7], [10]. Although 9 numeric attributes is not a large number, forming a non-linear separating hyperplane in SVM and determining nearest neighbors in KNN generally require adequate data density. With $n=120$ training samples, the spread of data points in a 9-dimensional space becomes relatively sparse (sparse data distribution), which can contribute to the high between-fold standard deviation in the cross-validation process.

The third factor is the difference in data type compared with previous studies. Studies that use wearable sensors [3] and electrocardiogram signals [4] measure physiological parameters directly and continuously, whereas this study uses questionnaire answers that are self-reported and measure stressors and perceived impact indirectly. Because data sources, populations, and evaluation protocols differ, performance values are not compared directly. Qualitatively, the limited performance in this study differs from the high accuracy (above 80%) reported for sensor-based approaches [3], [4], and the difference in data type is one possible explanation that requires further testing. Models based on academic questionnaires are therefore better regarded as an initial exploration or early risk screening rather than a precise diagnostic tool [8], [9].

The fourth factor is the grouping of the four ordinal sleep quality categories into two binary classes, which can place respondents in similar conditions into different classes, as well as possible measurement error in self-reported answers. In addition, other sleep quality factors not covered by the questionnaire may also play a role.

IV. CONCLUSION AND RECOMMENDATIONS

A. CONCLUSION

This study analyzed the association between academic assignment workload and students' sleep patterns and compared the performance of the SVM and KNN algorithms using a structured validation framework. The correlation analysis showed a moderate negative association ($r = -0.44$) between daily assignment work duration and sleep duration on active lecture days; this association is correlational and does not indicate a causal relationship.

Based on the primary evaluation with Nested Stratified 5-Fold Cross-Validation, the mean accuracy of SVM was 58.67% ($\pm 10.43\%$) and that of KNN was 51.33% ($\pm 7.67\%$), compared with a majority baseline of 57.33% ($\pm 1.49\%$). SVM was only 1.33 percentage points above the baseline and KNN was 6.00 percentage points below it, so the evidence that both models learn meaningful patterns remains limited. In the holdout test ($n=30$), SVM achieved an accuracy of 66.67% (ROC-AUC 0.7647) and KNN 56.67% (ROC-AUC 0.5724), equal to the baseline accuracy. Because the test set is small and the confidence intervals overlap, this difference is reported as supporting evidence and not as evidence of general superiority.

This study concludes that modeling sleep quality based on an academic assignment workload questionnaire yields limited predictive performance on this sample, which is possibly related to the variability of subjective perceptions between individuals, the small sample size, and class overlap. SVM showed higher mean values than KNN in both Nested CV and the holdout evaluation, but this difference is not strong enough to conclude that SVM is better in general.

B. ACADEMIC AND PRACTICAL IMPLICATIONS

Academically, this study contributes methodologically by showing the importance of applying Nested Cross-Validation and including a Baseline Classifier in machine learning modeling on perceptual data, in order to avoid claims of model superiority that are not supported by evidence (spurious performance). Practically, the association between longer assignment work duration and shorter sleep hours can serve as a consideration for study program administrators in designing more evenly distributed assignment deadlines, without implying a causal relationship.

C. LIMITATIONS AND FUTURE RESEARCH DIRECTIONS

This study has several limitations. The sample size is small ($n=150$), the composition of respondents follows per-faculty quotas and therefore does not represent the proportions of the student population, and the questionnaire data are self-reported and thus prone to individual perception bias. Extreme values were retained without a sensitivity analysis, the hyperparameter search space was limited to a particular grid, fold splits used fixed seeds with only five folds, and the performance of both models is close to the baseline. These results are therefore regarded as an initial exploration and not as a ready-to-use model. For future research, it is recommended to expand the sample size across campuses with a well-described sampling design, repeat cross-validation, conduct external validation, add features, integrate academic questionnaire data with objective biological sleep monitoring instruments (such as smartwatches or actigraphy), and explore non-linear feature selection techniques and ensemble learning algorithms (such as Random Forest or XGBoost) to improve classification stability and accuracy.

REFERENCES

- [1] S. Brisma dan D. B. Amalina, "Sleep disorders and their impact on brain health," *Nusantara Hasana Journal*, vol. 6, no. 1, pp. 100–107, Jun. 2026.
- [2] A. R. Putra, Alfarizi, dan Khikmawanto, "Pengaruh pola tidur terhadap kesehatan mahasiswa," *Jurnal Inovasi Global*, vol. 2, no. 9, pp. 1284–1290, Sep. 2024.
- [3] R. S. Pratama, I. Ramadhan, A. P. Wardhana, dan T. N. Abdillah, "Analisis literatur tentang penggunaan wearable IoT (smart watch, smart clothes, smart ring) untuk optimalisasi aktivitas olahraga," *Jurnal Penelitian Inovatif (JUPIN)*, vol. 5, no. 2, pp. 1857–1864, Mei 2025.
- [4] D. Yunita, F. S. Kusuma, A. Maulana, A. Zatayumni, Ihdaudin, dan D. B. Pradana, "Implementasi algoritma klasifikasi K-Nearest Neighbor (KNN) untuk klasifikasi level stres

- mahasiswa," *JRIIN: Jurnal Riset Informatika dan Inovasi*, vol. 2, no. 11, pp. 2045–2053, Apr. 2025.
- [5] M. A. A. Hilmi dan A. Susanto, "Implementasi dan evaluasi model machine learning untuk optimalisasi prediksi penjualan produk kue kering," *Building of Informatics, Technology and Science (BITS)*, vol. 7, no. 3, pp. 1649–1661, Des. 2025.
- [6] A. Bimantara T, "Deteksi berita palsu berbahasa Indonesia pada data tidak seimbang menggunakan IndoBERT dengan Focal Loss," Skripsi, Prog. Studi Teknik Informatika, Universitas Islam Negeri Maulana Malik Ibrahim, Malang, Indonesia, 2026.
- [7] S. Irviantina dan M. D. R. Siregar, "Optimalisasi splitting data untuk kinerja robust model EfficientNetV2-B0 pada deteksi pneumonia," *Majalah Ilmiah Methoda*, vol. 16, no. 2, pp. 83–89, Mei–Agu. 2026.
- [8] P. Tanamas, A. P. Ayu, A. L. Ekowati, dan E. Suryani, "Hubungan strategi koping dengan stres dan kualitas tidur pada mahasiswa kedokteran," *Damianus Journal of Medicine*, vol. 22, no. 3, pp. 232–242, Des. 2023.
- [9] A. Purwanto, "Pemodelan regresi multilevel pada pengaruh masalah yang dihadapi siswa terhadap hasil belajar matematika," *Jurnal Axioma: Jurnal Matematika dan Pembelajaran*, vol. 7, no. 2, pp. 119–128, Jul. 2022.
- [10] A. Sulistyawati dan D. W. Prabowo, "Analisis sentimen komentar pengguna Tiktok terhadap konten fashion brand Clotiva dengan Random Forest, Logistic Regression, dan Support Vector Machine," *JURAGAN: Jurnal Ragam Pengabdian dan Penelitian*, vol. 3, no. 1, pp. 2860–2872, Mar. 2026.