

Associations Among Social Media Use, Sleep Duration, and Mental Health Indicators in a Synthetic Adolescent Dataset

Ruixuan Zhao

The Bishop Strachan School, Toronto, Ontario M4V 1X2, Canada

How to cite this paper: Zhao, R. (2026). Associations among social media use, sleep duration, and mental health indicators in a synthetic adolescent dataset. *Social Sciences and Humanities*, 5(3), 62–73. ISSN Print: 3104-4328, ISSN Online: 3104-4336. <https://doi.org/10.63313/SSH.9144>
Published: 2026-10-22

Copyright © 2026 by author(s) and Erytis Publishing Limited.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
<http://creativecommons.org/licenses/by/4.0/>



Abstract

This study investigated the relationship between hours of social media using and hours of sleep duration, anxiety level, and stress level in a synthetic dataset containing 1,200 adolescent records. All four primary hypotheses were tested using Pearson correlations. A multiple linear regression analysis helped to examine whether daily social media use hours and screen time before bed were related to sleep duration, while accounting for age, gender, and physical activity. A one-way ANOVA compared sleep hours across four quartile-based social media use groups. The four primary correlations were approximately zero and statistically nonsignificant ($r_s = -.009$ to $.031$, all $p_s > .28$). The regression model explained almost none of the variance in sleep duration ($R^2 = .001$, $p = .895$), and average sleep duration did not differ significantly across the four usage groups ($p = .348$, $\eta^2 = .003$). The additional analyses of addiction and social interaction also found no significant relationships in this dataset. However, because the records were synthetically generated and the mental health measures were not documented as validated instruments, the results cannot be generalized to real adolescents. The study suggests the importance of evaluating dataset construction and measurement validity before using the findings to evaluate psychological theories.

Keywords

Social media use; Adolescent mental health; Sleep duration; Screen time; Synthetic data

1. Introduction

Social media is currently a major part of many teenagers' daily lives. Social media platforms such as TikTok and Instagram help teens communicate and maintain social connections, but researchers and healthcare professionals have started to raise concerns about possible links between social media use and teens' mental health (Weigle & Shafi, 2024), which often focus on sleep, stress, anxiety, and

depression, which may affect both teens' well-being and academic performance.

One proposed explanation is the displacement hypothesis, suggesting that time spent on social media may replace activities that support well-being, especially sleep. When teenagers use social media late at night, they may have less time to sleep. Since sleep supports emotional regulation and healthy development, reduced sleep may help explain the association between heavy social media use and poorer mental health (Weigle & Shafi, 2024).

This study examines whether patterns in a synthetic Kaggle dataset are consistent with these proposed relationships. By testing four theory-based associations within a single dataset, this study provides a direct empirical comparison with the patterns summarized in previous research. The dataset contains 1,200 adolescent records. The primary research question is: To what extent are daily social media hours associated with sleep duration, anxiety, and stress, and how is screen time before sleep associated with sleep duration? Using correlation analysis, multiple regression, and group comparisons, this study evaluates whether greater daily social media use and pre-bedtime screen use are associated with shorter sleep and whether daily social media use is associated with higher anxiety or stress.

2. Literature Review

Peer-related stress may also help explain why adolescents respond differently to social media. Two primary studies provide more specific evidence for these relationships. Woods and Scott (2016) surveyed 467 Scottish adolescents between the ages of 11 and 17. Overall social media use showed small associations with poorer sleep quality ($r = .24$) and higher anxiety ($r = .21$). Previous research suggests that the relationship between social media use and adolescent mental health is complex. Weigle and Shafi's (2024) literature review suggests that social media use generally has a small association with poorer youth mental health, although results vary across studies. The displacement hypothesis offers one possible explanation: screen time may replace other healthy activities, particularly sleep. Specifically, late-night social media use may delay bedtime or disrupt sleep, which may then affect emotional well-being. Moreover, the authors identify social comparison, fear of missing out, and problematic social media use as possible contributors to anxiety and depression. However, most existing studies are correlational and rely on self-reported data, making it difficult to establish whether social media directly causes these outcomes.

Two primary studies provide more specific evidence for these relationships. Woods and Scott (2016) surveyed 467 Scottish adolescents between the ages of 11 and 17. Overall social media use showed small associations with poorer sleep quality ($r = .24$) and higher anxiety ($r = .21$), while nighttime-specific use showed somewhat stronger associations with poorer sleep quality ($r = .34$) and anxiety ($r = .27$). Nighttime use also remained associated with poorer sleep after anxiety, depression,

and self-esteem were controlled ($\beta = .18$). Similarly, Li et al. (2019) analyzed data from 2,865 U.S. adolescents and found that social messaging was negatively associated with sleep duration ($\beta = -.149$). Their mediation analysis indicated that sleep characteristics fully mediated the association between social messaging and depressive symptoms. Together, these findings support examining nighttime screen use separately from total use and suggest that sleep may help explain the relationship between screen engagement and mental health.

Peer-related stress may also help explain why adolescents respond differently to social media. Zimmer-Gembeck and Skinner (2015) studied 669 adolescents through Grades 6 to 8 and examined their responses to peer-rejection threats using two short hypothetical scenarios. Adolescents with higher social anxiety reported stronger emotional reactions, more self-blame, and greater use of maladaptive coping strategies such as social avoidance and rumination. Those with more depressive symptoms also reported less perceived control and less use of adaptive coping strategies. Although this study did not examine social media, it provides evidence that responses to peer-rejection threats are associated with social anxiety, depressive symptoms, and coping during adolescence.

3. Hypotheses

Previous research suggest that social media use may affect adolescent well-being partly by displacing sleep. Weigle and Shafi (2024) reviewed research linking more time spent on social media, especially at night, to disrupted sleep and poorer mental health. Social comparison may also help explain the relationship between social media use and mental health. Social media frequently exposes adolescents to posts showing selected parts of their peers' appearance, achievements, and social lives. This may lead users to compare themselves negatively with others or experience fear of missing out. Constant peer feedback and the pressure to remain available online may also create psychological stress. Adolescents who spend more time on social media may experience these pressures more frequently, which may correspond with higher anxiety and stress levels (Weigle & Shafi, 2024).

H1: Daily social media hours will be negatively associated with sleep duration. As daily social media use increases, sleep duration is expected to decrease.

H2: Screen time before sleep will be negatively associated with sleep duration. Records with more screen use before bedtime are expected to show fewer hours of sleep.

H3: Daily social media hours will be positively associated with anxiety level. Greater social media use is expected to correspond with higher anxiety.

H4: Daily social media hours will be positively associated with stress level. Greater social media use is expected to correspond with higher stress.

Because H1 and H2 address related explanations for sleep duration, daily social media hours and screen time before sleep were also entered into the same multiple

regression model to examine whether either variable was associated with sleep duration after accounting for the other, age, gender, and physical activity. The association between anxiety and stress was retained as an exploratory consistency check rather than as a formal hypothesis.

4. Methodology

4.1. Research Design

This study used a quantitative, correlational secondary-data design to examine the associations between social media hours, sleep, and mental health indicators in a synthetic adolescent dataset. The study analyzed patterns already contained in the dataset and did not involve collecting new data. Because the design was correlational, the results were used to identify associations rather than causal relationships.

4.2. Data Source and Variables

The study used the public Teen Mental Health Dataset, downloaded from Kaggle (Kuasha, 2026). The dataset contained 1,200 synthetic adolescent records ($N = 1,200$), with each row representing one record. The gender variable labeled 615 records as male and 585 as female. Platform usage was divided into Instagram ($n = 411$), TikTok ($n = 398$), and both platforms ($n = 391$).

The dataset contained 13 variables: age, gender, daily social media hours, platform usage, sleep duration, screen time before sleep, academic performance, physical activity, social interaction level, stress level, anxiety level, addiction level, and a binary depression label. The primary variables examined in this study were daily social media hours, screen time before sleep, sleep duration, anxiety level, and stress level. The binary depression label was examined only in additional exploratory analyses.

Daily social media use ranged from 1.0 to 8.0 hours, screen time before sleep ranged from 0.5 to 3.0 hours, and sleep duration ranged from 4.0 to 9.0 hours. Anxiety, stress, and addiction were represented by scores ranging from 1 to 10, with higher scores indicating higher levels of each variable. The depression label was coded as either 0 or 1. Because the label was part of a synthetic dataset, it was treated as a constructed data variable rather than a clinical diagnosis or an independently verified mental health outcome. Analyses involving this label were therefore interpreted cautiously.

4.3. Data Preparation

Before analysis, missing values were checked across all 13 variables using Excel's COUNTBLANK function, and duplicate rows were checked across all 13 columns using Excel's Remove Duplicates tool. No missing values or duplicate rows were found. Therefore, all 1,200 records were retained, and the original data values

were not altered.

For the group analysis, records were divided into four quartile-based categories according to daily social media use: 1.0 to 2.8 hours ($n = 314$), greater than 2.8 to 4.5 hours ($n = 293$), greater than 4.5 to 6.3 hours ($n = 295$), and greater than 6.3 to 8.0 hours ($n = 298$). The cut points corresponded to the 25th, 50th, and 75th percentiles of daily social media use. Because usage hours were recorded to one decimal place and identical values at each boundary were kept in the same group, the four group sizes were not exactly equal.

4.4. Statistical Analysis

All analyses were conducted using Microsoft Excel. Descriptive statistics were calculated for all 13 variables. Means, standard deviations, minimums, and maximums were reported for numeric variables, while frequencies and percentages were reported for categorical variables. A correlation matrix was also created for the numeric and binary variables.

Pearson correlation coefficients (r) were calculated using Excel's CORREL function to measure the strength and direction of linear associations. The four primary correlations examined the following relationships:

1. Daily social media hours and sleep duration
2. Screen time before sleep and sleep duration
3. Daily social media hours and anxiety level
4. Daily social media hours and stress level

Two-tailed p -values were calculated by converting each correlation coefficient to a t statistic with $N - 2$ degrees of freedom. Ninety-five percent confidence intervals for the correlation coefficients were calculated using Fisher's r -to- z transformation. Because four primary correlation tests were conducted, a Bonferroni-adjusted significance level of .0125 was used for these tests. Exact, unadjusted p -values were also reported. Other inferential analyses used an alpha level of .05 and were identified as exploratory when they were not tests of the four primary hypotheses.

To examine H1 and H2 within the same model, a multiple linear regression was conducted with sleep duration as the dependent variable. Daily social media hours, screen time before sleep, age, gender, and physical activity were entered as predictors. The gender variable was dummy-coded as female = 1 and male = 0. The analysis reported unstandardized coefficients, standard errors, 95% confidence intervals, t statistics, p -values, R^2 , adjusted R^2 , and the overall F test. Variance inflation factors were examined to check for multicollinearity among the predictors.

A one-way analysis of variance (ANOVA) was used to compare mean sleep duration across the four quartile-based social media usage groups. Eta squared (η^2) was calculated as an effect-size measure. Ninety-five percent confidence intervals around each group mean were also calculated and displayed as error

bars in Figure 2. Comparisons across the Instagram, TikTok, and both-platform categories remained descriptive because platform differences were not part of the primary hypotheses.

Additional exploratory analyses examined the association between anxiety and stress, correlations involving addiction level, and differences across the three social interaction categories. Analyses involving the binary depression label were interpreted cautiously because the label was not an independently measured clinical diagnosis. Its correlations with sleep duration and daily social media hours were treated as point-biserial correlations. In addition, a Welch independent-samples *t* test compared sleep duration between records labeled 0 and 1, and Cohen's *d* was calculated to describe the standardized mean difference.

Scatterplots were visually inspected before the results were interpreted to identify obvious nonlinear patterns or extreme outliers. Results were interpreted using both *p*-values and effect sizes.

5. Results

5.1. Descriptive Statistics

All 1,200 synthetic records were retained for analysis. Table 1 summarizes all of the 13 variables in the dataset. Daily social media use averaged 4.54 hours per day (*SD* = 2.03), screen time before sleep averaged 1.74 hours (*SD* = 0.72), and sleep duration averaged 6.45 hours (*SD* = 1.44).

Table 1. Descriptive Statistics for All Dataset Variables.

Variable	Measurement	M (<i>SD</i>) or n (%)	Observed range or categories
Age	Years	15.93 (2.02)	13-19
Gender	Categorical	Male: 615 (51.3%); Female: 585 (48.8%)	Male, female
Daily social media use	Hours per day	4.54 (2.03)	1.0-8.0
Platform usage	Categorical	Instagram: 411 (34.3%); TikTok: 398 (33.2%); Both: 391 (32.6%)	Instagram, TikTok, both
Sleep duration	Hours per night	6.45 (1.44)	4.0-9.0
Screen time before sleep	Hours	1.74 (0.72)	0.5-3.0
Academic performance	Dataset score	2.99 (0.58)	2.0-4.0
Physical activity	Dataset score	1.01 (0.58)	0.0-2.0
Social interaction level	Categorical	Low: 415 (34.6%); Medium: 416 (34.7%); High: 369 (30.8%)	Low, medium, high
Stress level	1-10 score	5.45 (2.90)	1-10

Anxiety level	1-10 score	5.64 (2.86)	1-10
Addiction level	1-10 score	5.57 (2.83)	1-10
Depression label	Binary label	0: 1,169 (97.4%); 1: 31 (2.6%)	0, 1

Note. N = 1,200. Continuous variables are reported as M (SD); categorical variables are reported as n (%). The depression label is a synthetic dataset field and should not be interpreted as a clinical diagnosis.

5.2. Primary Correlation Analyses

Four Pearson correlations were used to test H1 through H4. Because four related primary tests were conducted, a Bonferroni-adjusted alpha level of .0125 (.05/4) was used; Table 2 reports the unadjusted two-tailed p values. None of the four associations was statistically significant, and every 95% confidence interval included zero. Therefore, H1 through H4 were not supported.

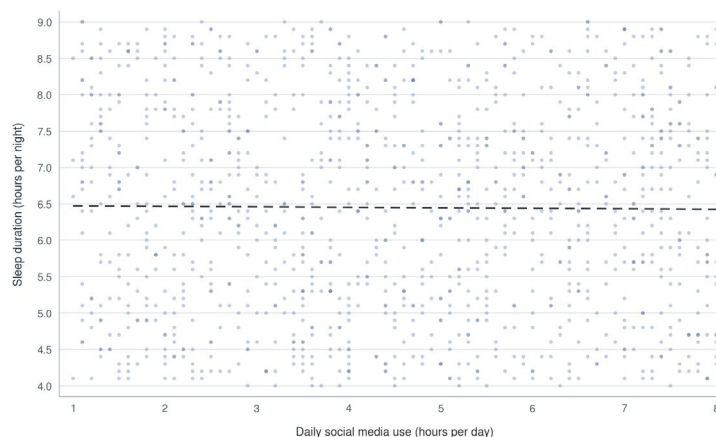
Table 2. Correlations Among Screen Use, Sleep, and Mental Health Variables

Hypothesis	Variables	R	p	Result	Result
H1	Daily social media hours and sleep duration	-.009	.743	[-.066, .047]	Not supported
H2	Screen time before sleep and sleep duration	.010	.723	[-.046, .067]	Not supported
H3	Daily social media hours and anxiety level	.028	.335	[-.029, .084]	Not supported
H4	Daily social media hours and stress level	.031	.288	[-.026, .087]	Not supported

Note. N = 1,200. Confidence intervals were calculated using Fisher's z transformation. The familywise alpha level was .0125; unadjusted two-tailed p values are shown.

Visual inspection of the scatterplots showed no clear upward or downward linear trends. Figure 1 illustrates the relationship between daily social media use and sleep duration. The small, evenly distributed points and nearly flat trendline are consistent with the near-zero correlation reported in Table 2.

Figure 1. Daily Social Media Use and Sleep Duration



Note. The dashed line represents the linear trendline. $r = -.009$, $p = .743$. $N = 1,200$.

Table 3. Correlations Among Numeric Dataset Variables

Variable	1	2	3	4	5	6	7	8	9	10
1. Age	—									
2. Daily social media use	-.007	—								
3. Sleep duration	.002	-.009	—							
4. Screen time before sleep	.076	.036	.010	—						
5. Academic performance	-.014	.013	.022	-.035	—					
6. Physical activity	.011	.026	.013	-.026	.023	—				
7. Stress level	-.031	.031	-.011	-.009	-.001	.012	—			
8. Anxiety level	.026	.028	-.012	-.010	-.064	-.022	.016	—		
9. Addiction level	.038	-.025	-.055	.029	.029	.026	.000	.031	—	
10. Depression label	.011	.175	-.191	-.017	.001	-.018	.170	.170	-.014	—

5.3. Multiple Regression Predicting Sleep Duration

A multiple regression tested whether daily social media use and screen time before sleep were associated with sleep duration after accounting for age, gender, and physical activity. The overall model was not statistically significant, $F(5, 1194) = 0.33$, $p = .895$, $R^2 = .001$, adjusted $R^2 = -.003$. Neither daily social media use nor screen time before sleep was a significant predictor, and the model explained approximately 0.1% of the variance in sleep duration. Variance inflation factors ranged from 1.001 to 1.009, indicating no evidence of problematic multicollinearity among the predictors.

Table 4. Multiple Regression Predicting Sleep Duration

Predictor	B	SE	t	p	95% CI for B
-----------	---	----	---	---	--------------

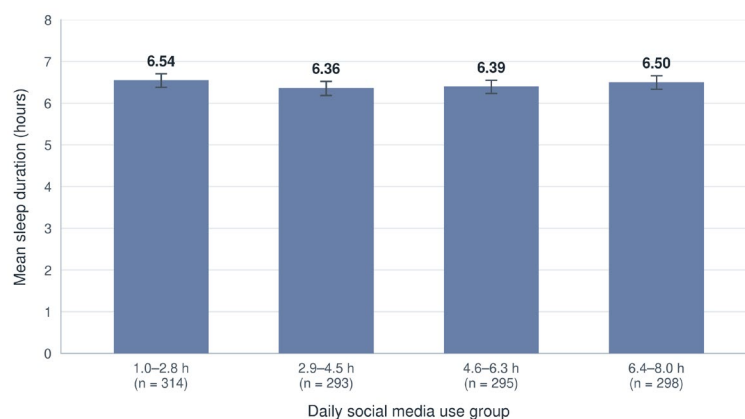
Intercept	6.357	0.362	17.567	< .001	[5.647, 7.067]
Daily social media use	-0.007	0.021	-0.339	.735	[-0.047, 0.033]
Screen time before sleep	0.020	0.058	0.346	.730	[-0.095, 0.135]
Age	0.001	0.021	0.031	.975	[-0.040, 0.041]
Female	0.091	0.083	1.094	.274	[-0.072, 0.255]
Physical activity	0.034	0.072	0.475	.635	[-0.107, 0.175]

Note. $N = 1,200$. The dependent variable was sleep duration in hours. Gender was coded male = 0 and female = 1.

5.4. Usage Group Analysis

Records were divided into four quartile-based groups using cut points of 2.8, 4.5, and 6.3 daily social media hours. Mean sleep duration ranged from 6.36 to 6.54 hours, a difference of approximately 0.19 hours between the highest and lowest group means. A one-way ANOVA found no statistically significant difference among the groups, $F(3, 1196) = 1.10$, $p = .348$, $\eta^2 = .003$. Because the omnibus test was not significant, no post-hoc comparisons were conducted.

Figure 2. Sleep Duration Across Daily Social Media Usage Groups



Note. Bars represent group means, and error bars show 95% confidence intervals. Values above the bars are mean sleep hours; sample sizes appear below the group labels. Equal values at quartile boundaries were kept in the same group. $N = 1,200$.

5.5. Exploratory Analyses

The association between anxiety and stress was examined as an exploratory consistency check rather than as a formal hypothesis. The correlation was near zero, $r = .016$, $p = .584$. Daily social media use was also not significantly associated with addiction level, $r = -.025$, $p = .388$.

Descriptive comparisons across social interaction categories showed little variation. Mean sleep duration was 6.46 hours in the low-interaction group, 6.41 hours in the medium-interaction group, and 6.48 hours in the high-interaction group. Mean

anxiety scores ranged from 5.53 to 5.70, mean stress scores ranged from 5.35 to 5.51, and mean addiction scores ranged from 5.47 to 5.64 across the three categories. These comparisons were descriptive and were not treated as tests of the primary hypotheses.

Finally, analyses involving the binary depression label were treated as exploratory. Sleep duration had a negative point-biserial correlation with the label, $r = -.191$, $p < .001$, while daily social media use had a positive point-biserial correlation with it, $r = .175$, $p < .001$. Records labeled 1 had lower mean sleep duration ($M = 4.76$, $SD = 0.57$, $n = 31$) than records labeled 0 ($M = 6.49$, $SD = 1.43$, $n = 1,169$). A Welch independent-samples t test showed a mean difference of -1.73 hours, $t(40.99) = -15.75$, $p < .001$, 95% CI $[-1.96, -1.51]$, Cohen's $d = -1.22$. Because the label is synthetic, highly imbalanced, and not documented as a clinical measure, this result describes the structure of the dataset rather than clinical depression among adolescents.

6. Discussion

The combined analyses did not show the sleep pattern predicted by the displacement hypothesis. Previous studies reported associations that were not observed in this dataset, although they used different measures of social media use and sleep. Woods and Scott (2016) reported correlations of $r = .24$ between overall social media use and poorer sleep quality and $r = .34$ between nighttime use and poorer sleep quality, while Li et al. (2019) reported a standardized association between social messaging and sleep duration ($\beta = -.149$). These comparisons should be interpreted cautiously because the previous studies used real participant data and different measures of social media use and sleep. However, these results do not disprove the displacement hypothesis. Weigle and Shafi (2024) explain that the effects of social media may depend on when, why, and how adolescents use it. Total daily hours and screen time before sleep may not distinguish between passive scrolling, active communication, and use that directly delays bedtime. In addition, the synthetic structure of the dataset may have weakened relationships that could appear in data collected from real adolescents.

The anxiety and stress findings may also reflect differences in measurement. Zimmer-Gembeck and Skinner (2015) examined specific factors such as peer rejection, social anxiety, emotional responses, and coping strategies. In contrast, the Kaggle dataset represents anxiety and stress using general scores from 1 to 10 and does not measure peer rejection, social comparison, or coping. Therefore, the near-zero associations in this study cannot directly confirm or challenge the psychological mechanisms examined in that research.

The exploratory analyses of addiction and social interaction did not reveal additional patterns. Daily social media use was not significantly associated with addiction level, and average sleep, anxiety, stress, and addiction scores differed only

slightly across the low, medium, and high social interaction categories. However, these variables were also represented by broad scales whose development and validity were not documented. As a result, the findings should not be interpreted as evidence that problematic social media use or social interaction is unrelated to adolescent well-being.

Records labeled 1 had substantially lower average sleep duration than records labeled 0, and the difference had a large, standardized effect size. Although the point-biserial correlation was relatively small, this partly reflects the severe group imbalance because only 31 of the 1,200 records were labeled 1, and the Kaggle documentation does not explain how the synthetic label was generated or establish it as a clinical measure of depression. The result therefore cannot be interpreted as evidence about clinically diagnosed depression. Overall, the study is more informative about the structure and limitations of this dataset than about real adolescent behavior.

7. Limitations

The biggest limitation is that the data were artificially generated and did not come from real adolescents. The Kaggle documentation does not fully explain how the records or relationships among variables were generated, so the near-zero associations may reflect the construction of the dataset rather than actual adolescent behavior. The dataset does not provide enough information about how anxiety, stress, addiction, and social interaction were measured. Similarly, the binary depression label is neither explained nor validated as a clinical measure. Also, the dataset only provides information at one point in time, so it cannot show whether changes in social media use are related to changes in sleep or mental health. Further analysis of this dataset could test nonlinear patterns and interaction effects among social media use, pre-bedtime screen use, addiction, and social interaction, as well as examine whether the findings remain stable under alternative grouping methods. However, conclusions about adolescent well-being would ultimately require real longitudinal data and validated psychological measures.

8. Conclusion

This study investigates whether the hours of using social media is associated with sleep duration hours, anxiety level, and stress level in a synthetic adolescent dataset. None of the analyses found significant relationships between social media using hours and the outcomes being studied. These findings do not demonstrate that social media has no effect on real adolescents because the dataset was artificially generated and its mental health variables were not documented as validated measures. The study also shows the importance of understanding the data before drawing conclusions from the results. Careful evaluation of dataset validity is therefore necessary before numerical results are used to support or challenge

theories about adolescent well-being.

References

- [1] Li, X., Buxton, O. M., Lee, S., Chang, A.-M., Berger, L. M., & Hale, L. (2019). Sleep mediates the association between adolescent screen time and depressive symptoms. *Sleep Medicine*, 57, 51–60. <https://doi.org/10.1016/j.sleep.2019.01.029>
- [2] Obaydul Kuasha. (2026, July 10). Teen_Mental_Health_Dataset. Kaggle datasets. <https://www.kaggle.com/datasets/obaydurrahmanrifat/teen-mental-health-dataset>
- [3] Weigle, P. E., & Shafi, R. M. A. (2024). Social Media and Youth Mental Health. *Current Psychiatry Reports*, 26(1), 1-8.
- [4] Woods, H. C., & Scott, H. (2016). #Sleepyteens: Social media use in adolescence is associated with poor sleep quality, anxiety, depression and low self-esteem. *Journal of Adolescence*, 51, 41–49. <https://doi.org/10.1016/j.adolescence.2016.05.008>
- [5] Zimmer-Gembeck, M. J., & Skinner, E. A. (2015). Adolescent vulnerability and the distress of rejection: Associations of adjustment problems and gender with control, emotions, and coping. *Journal of Adolescence*, 45, 149–159. <https://doi.org/10.1016/j.adolescence.2015.09.004>