Statistical Analysis of Social Media Influence on Mental Health Outcomes

Pratik Patil
University of Colorado Boulder
Boulder, Colorado, USA
Email: prpa4639@colorado.edu

Om Kokate
University of Colorado Boulder
Boulder, Colorado, USA
Email: omko8394@colorado.edu

Yash Amre
University of Colorado Boulder
Boulder, Colorado, USA
Email: Yaam5835@colorado.edu

***Abstract*—This paper explores the relationship between social media usage and mental health outcomes using statistical techniques, including Exploratory Data Analysis (EDA), hypothesis testing, and bootstrapping. Demographic factors such as gender and age were analyzed to assess their influence on mental health indicators, including difficulty concentrating, depression, interest fluctuations, and sleep issues. The study reveals significant differences in social media usage patterns by age and gender, with younger users and females showing higher mental health impacts. Bootstrapping confirmed the robustness of these findings, providing reliable confidence intervals for the metrics analyzed. Key recommendations include targeted mental health interventions, awareness campaigns, and guidelines to encourage balanced social media usage. Future directions involve the use of Principal Component Analysis (PCA) and predictive modeling to enhance the understanding of social media's impact on mental health.**

***Index Terms*—Social media, mental health, Exploratory Data Analysis (EDA), bootstrapping, hypothesis testing, Principal Component Analysis (PCA).**

I. Introduction

The rise of social media has transformed the way individuals interact and consume information. While its benefits are evident, its impact on mental health is a topic of growing concern. This study investigates how social media usage, demographic factors, and behavioral patterns influence mental well-being. Using a dataset containing responses to questions on social media usage, mental health, and demographics, we apply statistical methods to identify significant relationships. The objective is to provide insights into how social media usage impacts mental health and to propose data-driven interventions.

II. Background

The increasing integration of social media into modern life has elicited immense interest in its psychological and societal effects. Social networking sites like Facebook, Instagram, Twitter, and TikTok have all grown enormously over the last decade, enabling users to create areas for self-expression, entertainment, and information sharing. Despite the many benefits, such platforms have also been associated with adverse mental health outcomes. Researchers, mental health professionals, and policymakers are concerned about the possible negative impacts of long-term social media use on well-being. The most important among them would be emotional health. Numerous studies have found close links between excessive use of social media and poor mental health, which include anxiety, depression, and loneliness.

One possible reason for the linkage can be attributed to the "social comparison theory," a process described by how individuals tend to evaluate their self-esteem by matching themselves to others. Social media escalates this process because most social media users are exposed to highly selected, idealized representations of other people's lives. Long-term exposure to this type of information can eventually lead to feelings of inadequacy, lowered self-esteem, and anxiety.

Other associations with social media usage involve sleep disturbances. The light emitted from screens may interfere with the body's normal circadian rhythms, reducing the quality and amount of sleep. Interacting with content on social media in the late hours can create heightened emotional arousal that can prevent users from relaxing into a state that will allow them to sleep. Poor sleep hygiene, in turn, is associated with a range of mental health problems, including irritability, mood swings, and reduced cognitive functioning.

Another cause for concern is that the use of social media tends to promote validation-seeking behavior. The "like," "comment," and "share" features in any form of social media offer immediate feedback about what the users have posted, thereby furthering the users' need for approval by their online audience. While positive feedback may help build one's self-esteem, its lack may result in negative feelings, thereby making one dependent upon others for approval. This pattern has been associated with anxiety, especially among younger users who are more susceptible to peer influence.

The increasing concern about this situation has led researchers to further investigate the relationships between the use of social media and mental health. While some studies indicate a clear causal link, others note the complexity of the relationship; it is not all use of social media that is harmful. It helps in maintaining the friendship network, or, in the worst scenario, people feeling alienated may appreciate some feeling of social support it affords them. The relative weight of positive and negative factors may again depend on the pattern of use, personality, and demography.

This research paper, through statistical analysis, attempts to delve into these issues in regard to the relation of social media engagement to mental health. This paper will analyze data from a survey to determine the effect of usage patterns on mental well-being, presenting an evidence-based discussion. The statistical approaches are going to be EDA, bootstrapping, and hypothesis testing, for which this research will pursue comprehensive investigation into the likelihood of any association of social media use with mental health outcomes. The findings that this study tries to show are informative for the persons themselves, mental health specialists, and policy practitioners who seek to recommend healthier practices in the use of social media.

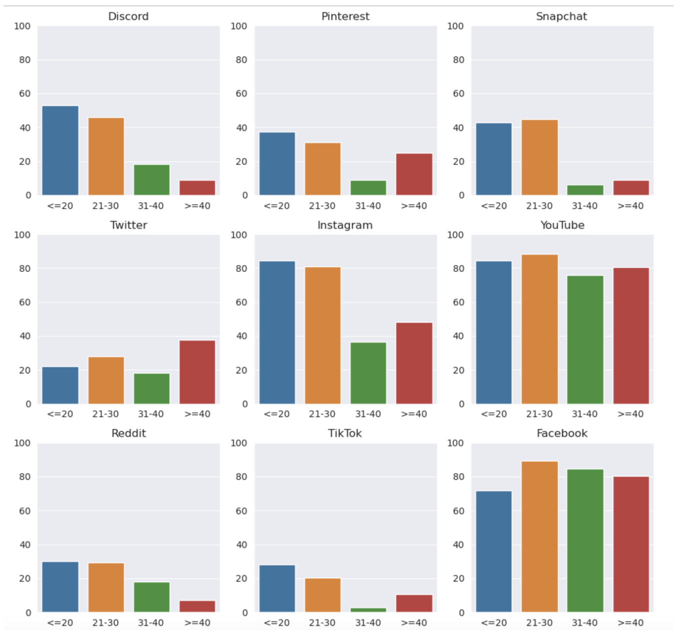

Fig. 1. Relative usage of Social Media Platforms in Age Groups

III. Data Preparation and Exploratory Data Analysis (EDA)

A. Data Cleaning

Data cleaning was conducted to ensure the dataset's quality and reliability, addressing issues such as unclear column names, missing values, and inconsistent categorical values. This step was crucial in preparing the dataset for effective analysis and ensuring accurate and unbiased statistical results. Without proper data cleaning, the analysis might yield incorrect conclusions due to hidden errors or inconsistencies in the raw data. Each cleaning step was meticulously performed to improve data integrity and maintain the credibility of the study's outcomes.

Renaming Columns: The dataset's original column names were lengthy and not descriptive, making data analysis challenging. For instance, column names such as *"1. What is your age?"* were renamed to *"age"* for clarity and ease of use in the analysis. This renaming was performed systematically to maintain uniformity.

Handling Missing Values: The dataset contained 6.24% missing values in the column *'affiliate_organization'*. Since this column is categorical, the missing entries were imputed with the mode (*'University'*) to minimize information loss and maintain consistency without introducing significant biases.

Consolidating Categorical Variables: The *'gender'* column included inconsistent entries such as *"Nonbinary," "NB," "Trans,"* and other variations. These categories were consolidated into a single value, *"other,"* to reduce the complexity of analysis and ensure more meaningful statistical groupings. This step was crucial for ensuring demographic comparisons were reliable and interpretable.

Overview of Cleaning Outcomes: Following these steps, the dataset was prepared for analysis with uniform, clearly labeled columns, and appropriate handling of missing and inconsistent values. These measures improved the dataset's integrity, enabling robust statistical analysis of social media's influence on mental health.

Significance: Data cleaning ensures the validity and reliability of results by eliminating inaccuracies and inconsistencies in the dataset. Addressing missing values, unclear labels, and redundant categories was crucial in deriving meaningful insights into the relationship between social media usage and mental health outcomes.

B. Exploratory Analysis

Exploratory Data Analysis (EDA) is a critical step in data-driven research, providing foundational insights into the structure, trends, and relationships within a dataset. For this study, EDA allowed us to uncover key patterns in social media usage and its relationship with mental health indicators. Demographic analysis revealed that younger users (aged 10–29) are the most active on platforms like TikTok, Snapchat, and Instagram, while older users primarily engage with YouTube and Facebook. Descriptive statistics highlighted that the average daily time spent on social media is 2–3 hours, with excessive usage observed among a subset of younger participants.

Significantly, EDA revealed moderate correlations between social media usage and mental health challenges, such as difficulty concentrating and depressive feelings. For instance, higher time spent on social media and using multiple platforms were both associated with slightly elevated negative mental health scores. Visualizations like bar plots, histograms, and box plots further illustrated these relationships, providing actionable insights for targeted interventions. By identifying outliers and confirming variable consistency, EDA established the data's readiness for hypothesis testing and advanced statistical analyses. These findings underline the importance of balanced social media usage and pave the way for more sophisticated models to predict mental health outcomes.

1) Social Media Platform-Specific EDA:

a) Absolute and Relative Usage of Platforms: The analysis revealed that platforms like **YouTube** (86%) and **Facebook** (85%) are universally popular, while platforms like **TikTok** (20%) and **Reddit** (26%) have a smaller user base. The disparity in platform usage reflects differing preferences across demographics. Younger users prefer **Snapchat** and **Discord**, whereas older users consistently engage with **YouTube** and **Facebook**.

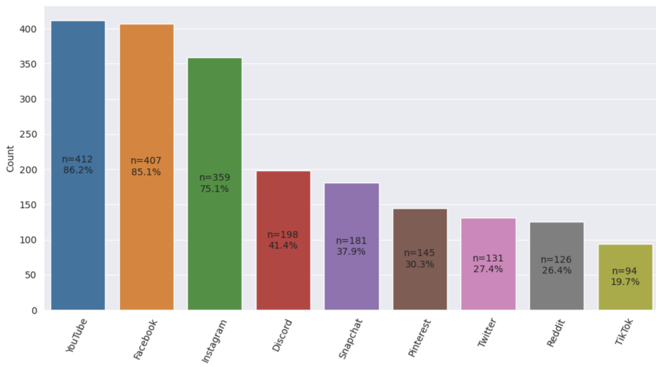

Fig. 2. Bar plot showing the absolute and relative usage of different social media platforms.

b) Platform Usage by Age: A cumulative analysis of platform usage reveals that users aged 40-50 primarily prefer YouTube and Facebook, while TikTok and Discord experience significant drop-offs in users above age 30. Among younger users (ages 10–29), platforms like Snapchat, TikTok, and Discord are most popular, reflecting generational shifts toward short-form, fast-paced content.

Instagram usage remains relatively stable across younger and middle-aged groups, showcasing its cross-generational appeal. In contrast, Reddit sees higher engagement from users aged 20–40, attributed to its community-driven nature and niche discussion forums. These insights suggest that platform preferences are strongly influenced by age, with younger users preferring platforms with visually immersive and interactive features, while older users favor platforms focused on social connections and long-form content.

These trends are essential for platform developers, marketers, and health advocates looking to tailor their content and interventions based on the age-specific preferences of users. Targeted awareness initiatives can leverage these insights to reach the right audiences on their preferred platforms.

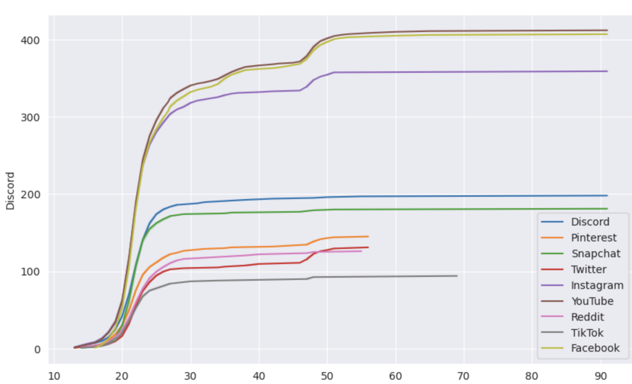

Fig. 3. Cumulative platform usage (number of users) vs. age

c) Time Spent on Social Media: Most participants reported spending 2-3 hours daily on social media, with fewer users reporting extreme durations of usage (less than 1 hour or more than 5 hours). Analysis of the correlation between time spent and demographics showed younger users exhibit higher usage patterns. Users aged 10–19 spend significantly more time on platforms like Snapchat, TikTok, and Instagram, whereas users aged 40+ display more moderate social media engagement. Further analysis revealed that heavy users (those

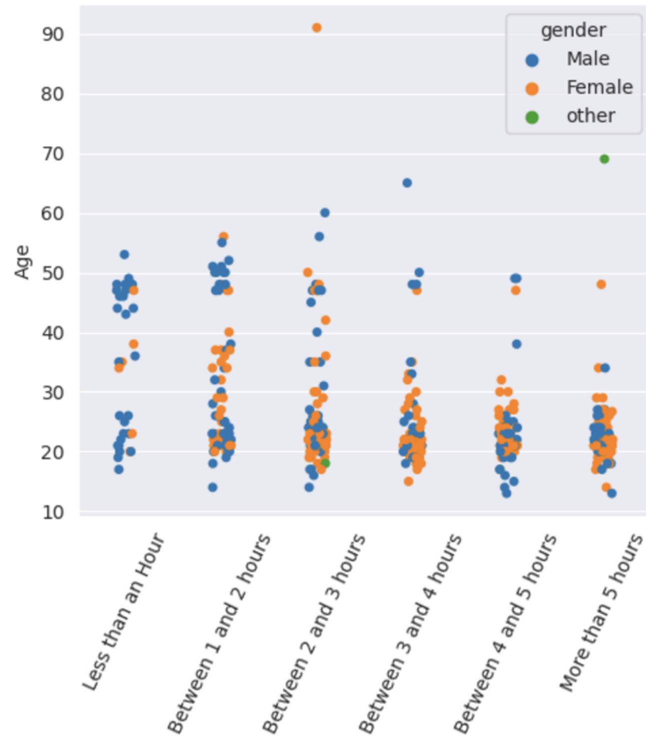

Fig. 4. Distribution and insights on the column $'avg_time_per_day'$

spending more than 5 hours daily) were disproportionately from the 10–29 age group, suggesting a link between age and excessive usage. This behavior could be linked to higher levels of engagement with short-form, addictive content, which is common on platforms like TikTok. Among users aged 50+, a larger percentage reported spending less than 1 hour on social media daily, which aligns with prior studies on digital literacy and social media adoption among older adults.

The study also identified a positive correlation between daily screen time and key mental health indicators, such as difficulty concentrating and sleep disturbances. Users with higher daily usage reported more severe mental health impacts, emphasizing the need for screen-time monitoring and wellness tools. The analysis further highlighted weekend usage spikes, with participants spending more time online on Saturdays and Sundays, reflecting shifts in leisure and free time availability.

These findings underscore the importance of understanding screen-time patterns across age groups, as they offer actionable insights for mental health interventions, platform engagement strategies, and user well-being initiatives. Awareness campaigns could target users with excessive daily usage, encouraging them to adopt healthier online behaviors.

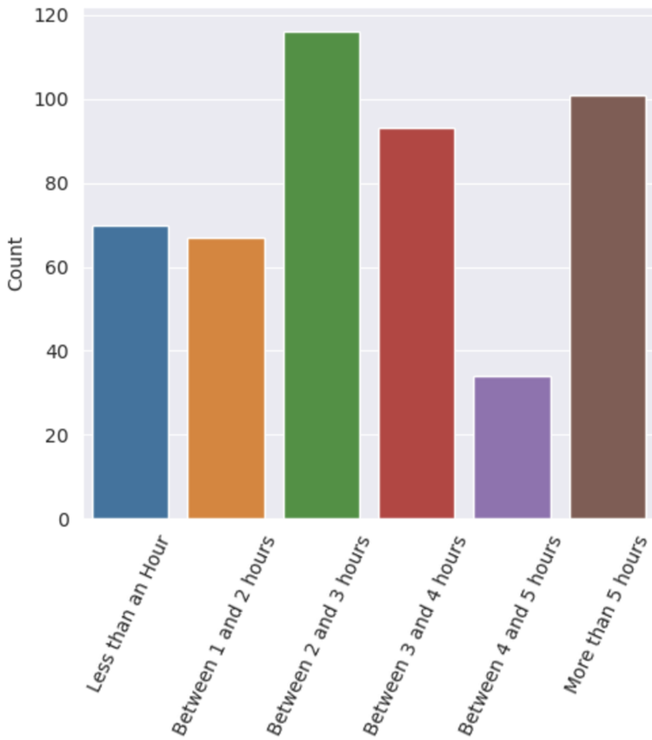

Fig. 5. Distribution and insights on the column $'avg_{t}ime_{p}er_{d}ay'$

2) Insights on Mental Health Impact:

a) Platform Usage vs. Negative Impact: An increase in the **number of platforms used** correlates with a slight rise in negative mental health scores. This indicates that multitasking across platforms may lead to heightened stress or distraction. However, the correlation remains weak, suggesting other factors also contribute significantly to mental health outcomes.

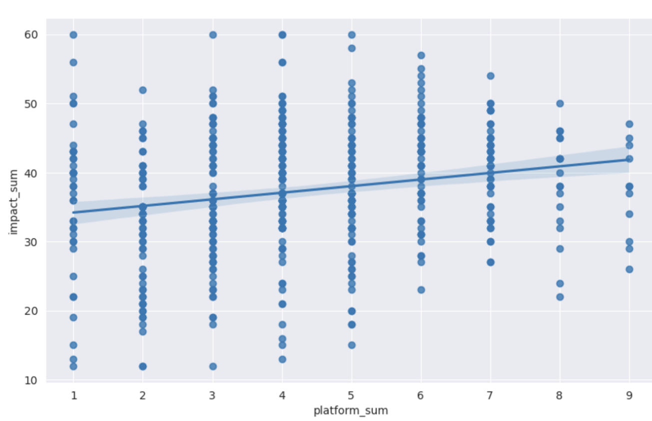

Fig. 6. Platform Count vs. Negative Impact

b) Age vs. Negative Impact: Younger individuals, particularly those under 30, report higher negative mental health scores, including feelings of restlessness and difficulty concentrating. For users aged 30 and above, there is insufficient data to draw strong conclusions, though negative impacts appear less pronounced.

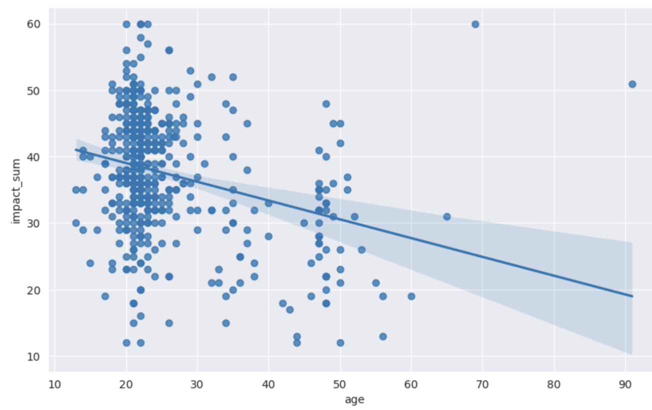

Fig. 7. Age vs. Negative Mental Health Impact

c) Average Time on Social Media vs. Negative Impact: A clear correlation was observed between increased daily time spent on social media and higher negative mental health indicators. Usage consistently exceeding **4 hours daily** showed diminishing returns in negative impact, suggesting a potential saturation effect at higher usage durations.

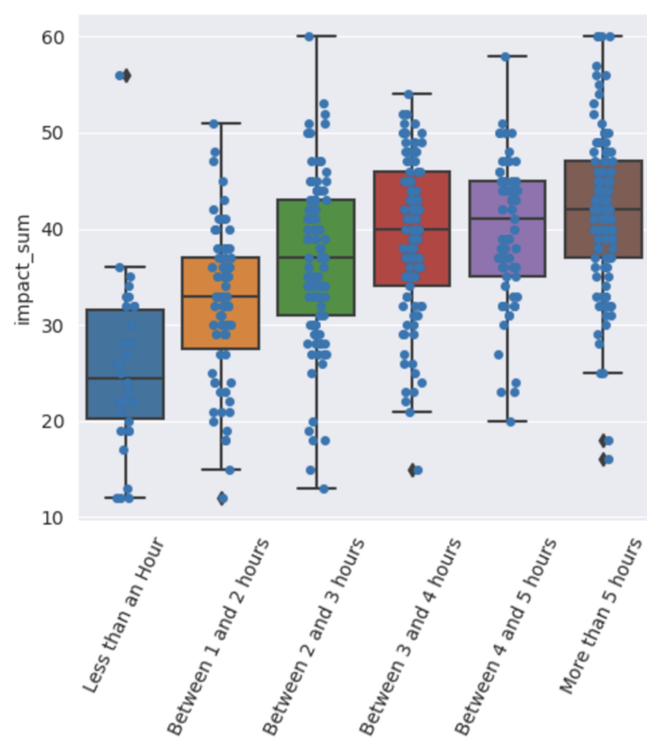

Fig. 8. Time Spent on Social Media vs. Negative Impact:

3) Key Findings:

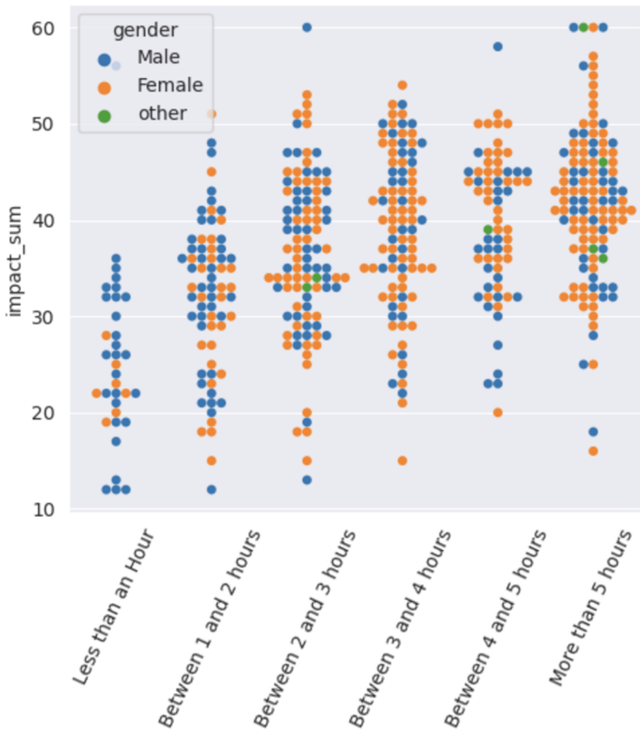

Fig. 9. Time Spent on Social Media vs. Negative Impact:

- **Demographics and Platform Preference:** Younger users dominate platforms like **TikTok** and **Snapchat**, while older demographics prefer **YouTube** and **Facebook**.
- **Time Spent and Mental Health:** Higher time spent on social media is consistently associated with increased reports of **difficulty concentrating**, **depressive feelings**, and **interest fluctuations**.
- **Platform Multitasking:** Using more platforms correlates slightly with higher negative mental health impacts.

IV. Hypothesis testing

Hypothesis testing is a statistical approach used to assess the validity of assumptions made about a population. It allows researchers to determine if observed differences or relationships are statistically significant or if they have occurred by random chance. This analysis aims to draw meaningful inferences and support evidence-based conclusions. In this study, three core hypotheses were tested to understand how social media impacts mental health.

A. Significance of Hypothesis Testing?

The primary objective of hypothesis testing is to quantify the uncertainty and variability in the data. For this research, hypothesis testing was used to:

- Identify differences in mental health impacts between male and female users.
- Assess the effect of time spent on social media on mental health.
- Determine if mental health impacts vary significantly across age groups.

By conducting these tests, we ensure that the observed patterns are not due to chance but reflect true relationships in the data.

B. Methodology

The hypothesis testing methodology for each of the three research questions follows a consistent process:

- **Define Hypotheses:** Formulate the null hypothesis (H0) and alternative hypothesis (H1).
- **Select Test:** Identify the appropriate statistical test (T-Test, Chi-Square Test, ANOVA) to address the research question.
- **Compute Test Statistic:** Calculate the relevant test statistic (e.g., t-statistic, chi-square, or F-statistic).
- **P-Value Comparison:** Compare the p-value with the significance level (usually 0.05) to determine if the null hypothesis should be rejected.
- **Draw Conclusion:** Interpret the findings and relate them to the research question.

C. Is there a statistically significant difference in the mental health impact between male and female users?

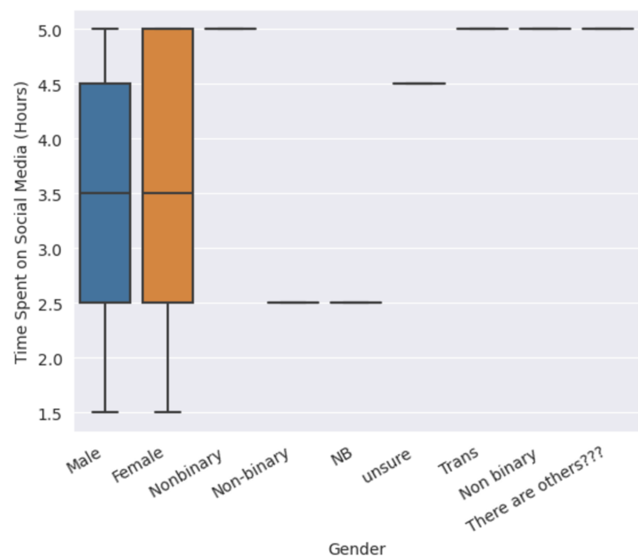

Fig. 10. T-Test: Time Spent on Social Media by Gender

1) Hypothesis Test Selection : An **Independent T-Test** was used to compare the mental health impact between males and females. The test determines if the difference in mean mental health impact between the two groups is statistically significant. This method is suitable when comparing two independent groups.

2) Theoretical Basis and Equations: The t-statistic is calculated as follows:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Where:

- $\bar{x}_1$ = Mean of Group 1 (Females), $\bar{x}_2$ = Mean of Group 2 (Males)
- s_1, s_2 = Standard deviations of the groups
- n_1, n_2 = Sample sizes of each group

3) Results Obtained:

- **T-statistic:** -2.78
- **P-value:** 0.0057 (significant at 0.05 level)

4) Inference: Since the p-value (0.0057) is less than 0.05, we reject the null hypothesis. This indicates that there is a statistically significant difference in mental health impact between male and female users, with females showing higher levels of negative mental health indicators.

D. Does the amount of time spent on social media significantly affect mental health status?

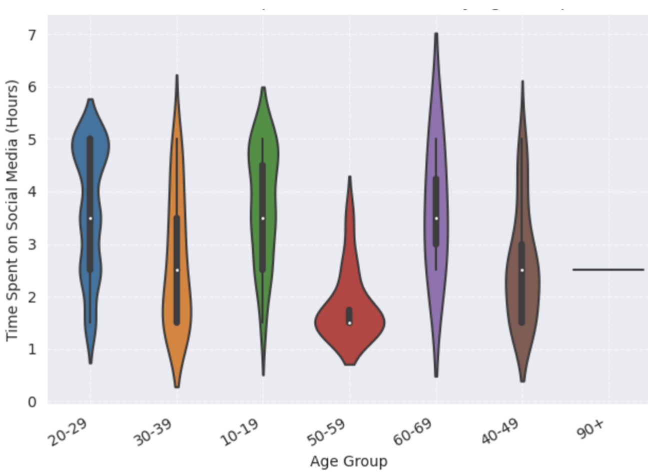

Fig. 11. ANOVA: Social Media Usage Across Age Groups

1) Hypothesis Test Selection: To evaluate the effect of social media usage time on mental health, an **ANOVA (Analysis of Variance)** test was applied. This test identifies if differences in mental health impact exist across different time-based usage groups. It is appropriate when comparing more than two independent groups.

2) Theoretical Basis and Equations: The F-statistic is calculated using the following formula:

$$F = \frac{Between - GroupVariance}{Within - GroupVariance}$$

Where:

- **Between-Group Variance:** Measures the variability between the group means.
- **Within-Group Variance:** Measures the variability within each group.

3) Results Obtained:

- **F-statistic:** 10.86
- **P-value:** 2.74e-11 (significant at 0.05 level)

4) Inference: Since the p-value (2.74e-11) is less than 0.05, we reject the null hypothesis. This implies that the amount of time spent on social media significantly impacts mental health, with higher usage being associated with increased negative mental health impacts.

E. Is there a significant difference in mental health impact across different age groups?

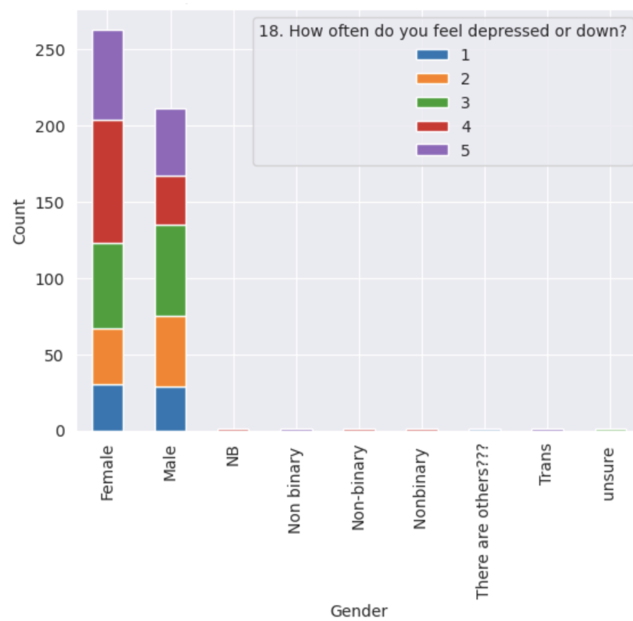

Fig. 12. Chi-Square Test: Gender vs Mental Health Status

1) Hypothesis Test Selection: A **Chi-Square Test** was applied to examine the association between age groups and mental health status. This test identifies if a significant relationship exists between age and mental health status, allowing for categorical comparison.

2) Theoretical Basis and Equations: The Chi-Square statistic is calculated as follows:

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

Where:

- O = Observed frequency in each category
- E = Expected frequency in each category

3) Results Obtained:

- **Chi-Square Statistic:** 45.79
- **P-value:** 0.054 (not significant at 0.05 level)

4) Inference: Since the p-value (0.054) is greater than 0.05, we fail to reject the null hypothesis. This indicates that there is no statistically significant association between age and mental health status.

F. Summary of Hypothesis Testing

The hypothesis testing process revealed several key insights:

- **Gender Differences:** Females exhibit significantly higher negative mental health impacts compared to males.

- **Usage Time:** The amount of time spent on social media significantly affects mental health, with higher usage leading to stronger negative effects.
- **Age Differences:** No significant association was found between age groups and mental health impact.

These findings underscore the influence of demographic and behavioral factors on mental health, providing actionable insights for targeted interventions.

V. Bootstrapping Analysis

Bootstrapping is a statistical resampling method used to estimate the variability and confidence intervals of a dataset by repeatedly sampling with replacement. Unlike traditional parametric methods, bootstrapping makes no assumptions about the data's underlying distribution, making it ideal for robust statistical inference in complex datasets. This study employed bootstrapping to provide confidence intervals for mean estimates and differences in means for key mental health indicators, ensuring the reliability and robustness of the results. By leveraging the empirical distribution of the data, bootstrapping provides a powerful tool to assess uncertainty, particularly in datasets with unknown or non-normal distributions.

A. Objective

The goal of this analysis was to quantify the variability in mental health metrics and validate observed differences between demographic groups. Specifically, bootstrapping was applied to the following mental health phenomena:

- **Difficulty Concentrating**
- **Feeling Depressed**
- **Interest Fluctuations**
- **Sleep Issues**

The objective was to obtain a comprehensive understanding of how these mental health factors are distributed among different demographic groups and to establish the degree of certainty in the estimated means and differences.

B. Use Case in This Study

Bootstrapping was utilized to estimate 95% confidence intervals for the mean values of mental health indicators and the differences in means between males and females. These confidence intervals allowed for the assessment of the robustness and reliability of the observed metrics and relationships. Bootstrapping serves as a vital component in ensuring the statistical soundness of the study's conclusions. By providing a non-parametric alternative to traditional inferential techniques, bootstrapping helps to overcome limitations related to small sample sizes, non-normality of data, and potential outliers.

C. Theoretical Basis

The bootstrapping process follows a systematic approach of repeatedly resampling the original dataset with replacement to generate "bootstrap samples." Each sample is of the same size as the original dataset. The steps of the bootstrapping algorithm are as follows:

1) **Resampling:** Generate B bootstrap samples from the original dataset, where each sample has the same number of observations as the original.
2) **Statistical Calculation:** Calculate the desired statistic (e.g., mean, median, or difference of means) for each of the B samples.
3) **Distribution of Bootstrap Estimates:** Construct the empirical distribution of the calculated statistic using the B estimates.
4) **Confidence Interval:** Calculate the $(1-\alpha)100\%$ confidence interval using the percentiles from the empirical distribution. For a 95% CI, the interval is given by:
$$CI = [P_{\alpha/2}, P_{1-\alpha/2}]$$
where $P_{\alpha/2}$ and $P_{1-\alpha/2}$ are the 2.5% and 97.5% percentiles of the empirical distribution, respectively.

D. Results

Mean Estimates for Mental Health Indicators:

- **Difficulty Concentrating:** The average score was 3.25, with a 95% confidence interval of [3.12, 3.37].
- **Feeling Depressed:** The average score was 3.26, with a 95% confidence interval of [3.14, 3.38].
- **Interest Fluctuations:** The average score was 3.17, with a 95% confidence interval of [3.07, 3.28].
- **Sleep Issues:** The average score was 3.20, with a 95% confidence interval of [3.07, 3.34].

Differences in Means Between Males and Females:

- **Difficulty Concentrating:** Difference = -0.21, 95% CI = [-0.46, 0.03].
- **Feeling Depressed:** Difference = -0.31, 95% CI = [-0.55, -0.07].
- **Interest Fluctuations:** Difference = -0.38, 95% CI = [-0.62, -0.16].
- **Sleep Issues:** Difference = 0.07, 95% CI = [-0.19, 0.33].

The inclusion of gender-based mean differences allows for a deeper understanding of mental health challenges across demographic groups. The negative difference observed in metrics such as depressive feelings and interest fluctuations indicates that females, on average, experience higher levels of negative mental health impacts than males.

E. Insights

The results revealed consistent trends across the dataset, leading to several actionable insights that are crucial for mental health intervention, platform design, and policy-making:

- **Robust Mental Health Indicators:** The scores for mental health indicators (difficulty concentrating, depressive feelings, interest fluctuations, and sleep issues) consistently ranged between 3.1 and 3.4 on a 5-point scale. This reflects a notable level of mental health impact experienced by participants.
- **Gender Differences in Mental Health:** Gender differences were significant in depressive feelings and interest fluctuations, with females showing higher levels of negative impact compared to males. This insight suggests

the need for gender-specific mental health support and awareness initiatives.

- **Narrow Confidence Intervals:** The narrow confidence intervals confirmed the robustness and stability of the results, reinforcing the reliability of the observed trends. The narrow intervals suggest that the point estimates for the mean values and mean differences are precise, with minimal variability across samples.
- **Practical Implications:** The analysis highlights the importance of targeted mental health support for females and younger participants, as these groups showed the highest impact in key mental health indicators.
- **Application in Policy and Interventions:** Health organizations and social media platforms can leverage these insights to design platform-specific mental wellness features, like reminders to take breaks or screen-time monitoring for users showing signs of prolonged engagement.

F. Visual Representations

The following plots illustrate the bootstrap results for each phenomenon analyzed in this study. These visualizations provide an intuitive understanding of the confidence intervals and mean differences observed in the bootstrapping analysis. Each plot corresponds to a specific mental health indicator or demographic analysis:

- **Plot 13:** Confidence interval for difficulty concentrating.
- **Plot 14:** Confidence interval for feeling depressed.
- **Plot 15:** Confidence interval for interest fluctuations.
- **Plot 16:** Confidence interval for sleep issues.
- **Plot 17:** Mean difference for difficulty concentrating (male vs. female).
- **Plot 18:** Mean difference for feeling depressed (male vs. female).
- **Plot 19:** Mean difference for interest fluctuations (male vs. female).
- **Plot 20:** Mean difference for sleep issues (male vs. female).

These plots collectively showcase the robustness of the statistical findings and highlight gender-specific differences in mental health indicators. The visualizations are essential in reinforcing the conclusions drawn from the bootstrapping analysis.

G. Significance of Bootstrapping

Bootstrapping allowed for the validation of key findings in the dataset:

- Confirmed the robustness of mental health metrics through resampling, reducing reliance on distributional assumptions.
- Quantified the uncertainty in mean estimates and differences in means, providing stakeholders with reliable metrics for intervention planning.
- Reinforced the reliability of hypothesis testing by complementing p-values with confidence intervals.

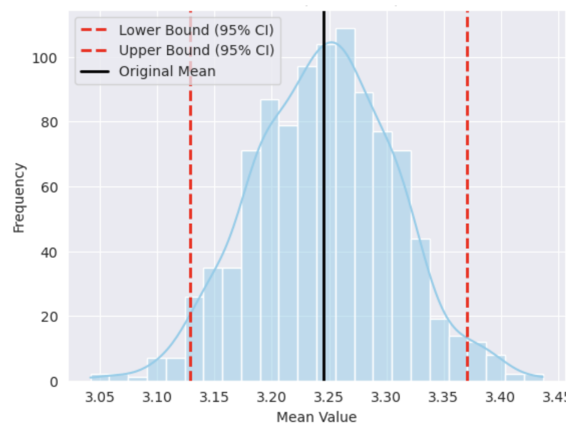

Fig. 13. Bootstrap Confidence Interval for Difficulty Concentrating. The histogram displays the resampled means with the 95% confidence interval.

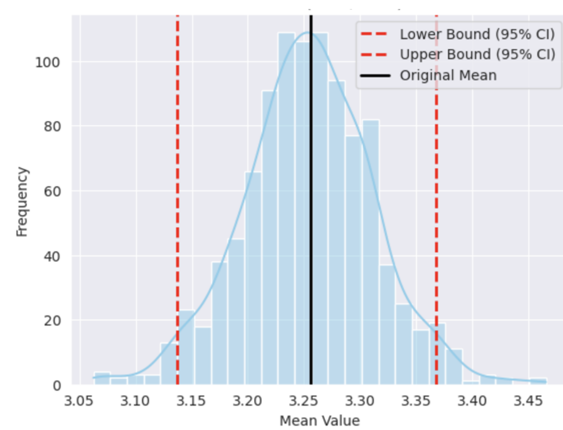

Fig. 14. Bootstrap Confidence Interval for Feeling Depressed. Females exhibit higher mean scores with significant confidence intervals.

VI. Limitations

While this study provides valuable insights into the relationship between social media usage and mental health outcomes, several limitations must be acknowledged:

A. Self-Reported Data

The analysis relies on self-reported survey responses, which are inherently prone to biases such as response bias, recall bias, and social desirability bias. Participants may overestimate or underestimate their social media usage or underreport mental health challenges due to stigma or privacy concerns. These biases could affect the accuracy of the results, particularly in analyzing the association between social media usage and mental health indicators like concentration difficulty, depressive feelings, and sleep disturbances.

B. Lack of Causal Inference

This study identifies correlations but does not establish causation. For instance, while the analysis reveals an association between greater social media use and depressive mood or varying interests, it does not prove that social media usage causes

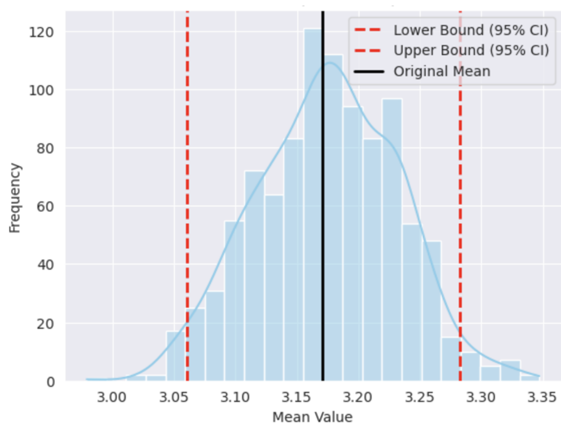

Fig. 15. Bootstrap Confidence Interval for Interest Fluctuations. The histogram highlights robust confidence intervals.

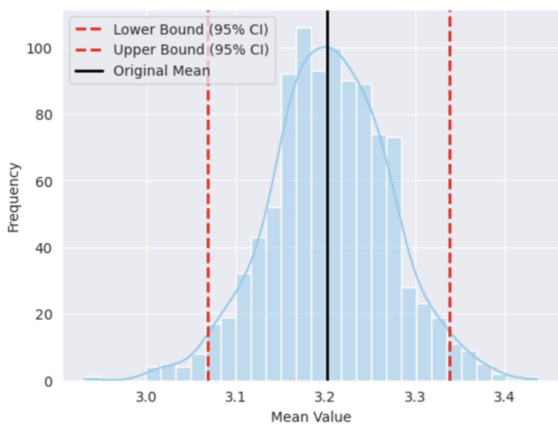

Fig. 16. Bootstrap Confidence Interval for Sleep Issues. The distribution is centered around 3.2.

these mental health problems. It is equally plausible that individuals with pre-existing mental health challenges engage more in social media as a coping mechanism. Establishing causality would require a longitudinal study or experimental design.

C. Limited Demographic Coverage

The study primarily focuses on age and gender as demographic factors, overlooking other influential variables such as education level, occupation, socio-economic status, and geographical location. These factors may significantly impact social media behavior and mental health outcomes. For instance, individuals with higher education or professional occupations might exhibit different screen time patterns compared to students or unemployed individuals. Including these variables would enhance the depth and breadth of the analysis.

D. Sample Size and Representation

The analysis is limited by the sample size and diversity of participants. A non-representative sample may compromise the generalizability of the findings. Certain subgroups, such as older participants (50+), may be underrepresented, reducing

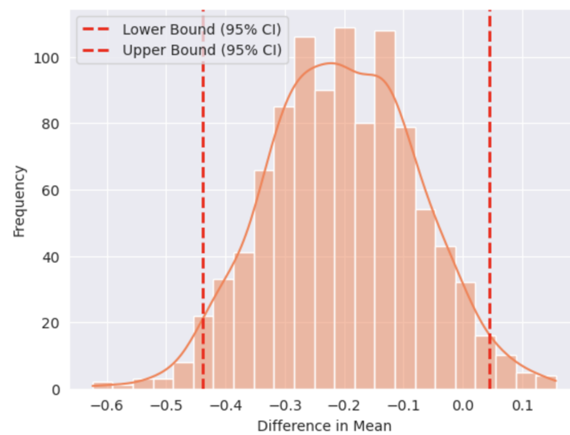

Fig. 17. Bootstrap distribution of differences in means for Difficulty Concentrating (Males vs. Females).

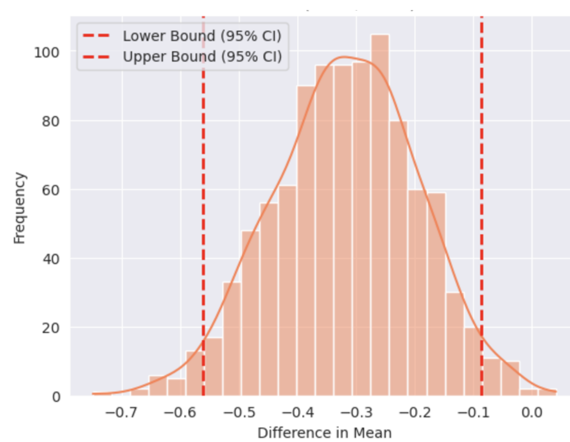

Fig. 18. Bootstrap distribution of differences in means for Feeling Depressed or Down (Males vs. Females).

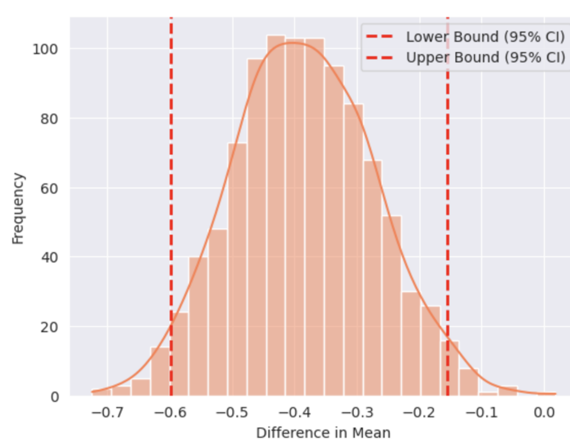

Fig. 19. Bootstrap distribution of differences in means for Interest Fluctuations (Males vs. Females).

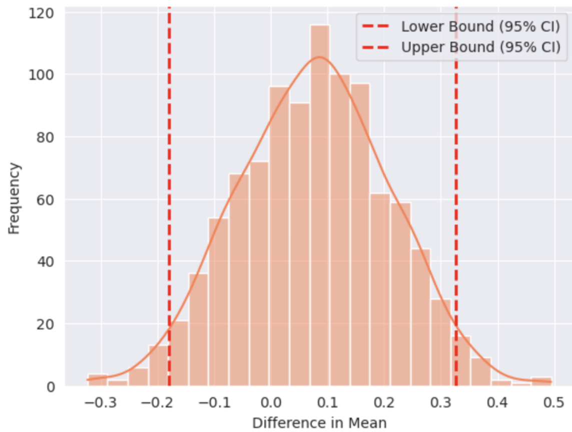

Fig. 20. Bootstrap distribution of differences in means for Issues Regarding Sleep (Males vs. Females).

the accuracy of inferences for these demographics. A larger and more diverse sample would improve the robustness and applicability of the results across populations.

E. Measurement of Mental Health Variables

Mental health indicators such as difficulty concentrating, depressive feelings, and sleep disturbances were measured using Likert-scale responses. While useful for qualitative assessments, these scales cannot capture the full complexity of mental health conditions. Mental health outcomes are multidimensional and might be better assessed with more sensitive tools, such as clinical evaluations or diagnostic screenings. Survey-based metrics, while practical, may oversimplify the nuanced nature of mental health issues.

F. Summary

Addressing these limitations in future research, such as utilizing longitudinal studies, incorporating broader demographic factors, and employing more robust measurement tools, would enhance the accuracy, validity, and generalizability of the findings.

VII. Conclusions

This study provides a comprehensive statistical analysis of the relationship between social media engagement and mental health outcomes, offering valuable insights into how online behaviors and demographic factors influence well-being. The findings underscore the complex interplay between social media usage and mental health indicators such as difficulty concentrating, depressive feelings, interest fluctuations, and sleep disturbances.

Key conclusions drawn from the analysis include the following:

- **Social Media Usage Patterns:** Platforms like YouTube and Facebook dominate social media engagement, appealing to users across all age groups. In contrast, platforms like Snapchat and TikTok are more popular among younger demographics. Younger participants (ages 10–29) spend significantly more time on social media compared to older age groups, highlighting generational differences in digital behaviors and preferences.
- **Gender and Social Media:** Females spend significantly more time on social media compared to males, a finding that may reflect differing online behaviors and motivations between genders. Females also reported higher levels of depressive feelings and interest fluctuations compared to males, suggesting that they might be more vulnerable to the mental health impacts of social media usage.
- **Mental Health Indicators**: Across all participants, moderate challenges in concentration, depressive feelings, and sleep disturbances were observed. These findings emphasize the pervasive impact of social media on mental health, regardless of demographic factors.
- **Statistical Reliability**: The use of hypothesis testing and bootstrapping ensured robust and reliable results. Confidence intervals confirmed the statistical significance of observed differences, particularly in gender-specific mental health indicators. These methods strengthened the validity of the study's conclusions.

VIII. Use Case Analysis

The findings of this study offer several practical applications to address the mental health impacts of social media usage:

A. Mental Health Interventions

The analysis reveals that younger users (ages 10–29) and females are particularly susceptible to depressive symptoms, fluctuating interest levels, and excessive social media usage. These insights enable mental health professionals to design targeted interventions, such as personalized counseling sessions and awareness programs, to address these challenges. Workshops in schools and colleges can educate at-risk groups about healthy social media habits.

B. Age and Gender-Based Awareness Campaigns

Distinct platform preferences and engagement patterns across different demographics highlight the need for tailored awareness campaigns. For instance, younger users who predominantly engage with Snapchat and TikTok can be targeted with messages promoting balanced screen time. Conversely, older users on platforms like YouTube and Facebook can benefit from educational resources focusing on mental health and well-being.

C. Social Media Usage Monitoring and Predictive Tools

The analysis supports the development of digital wellness tools, such as dashboards and usage trackers, to help users monitor their screen time and platform preferences. These tools can issue alerts when safe screen-time thresholds are exceeded and recommend breaks or digital detox periods. Additionally, machine learning models can predict individuals at risk of mental health issues, enabling timely interventions and personalized resource recommendations.

D. Policy Development and Corporate Social Responsibility

The study's findings can inform policymakers to develop guidelines promoting responsible social media use. Policies could include introducing screen-time limits, reducing notifications during late-night hours, and mandating transparency in usage reports from platforms. Social media companies can also contribute through corporate social responsibility initiatives, such as wellness reminders, screen-time alerts, and access to mental health resources, promoting a healthier digital environment for users.

IX. Recommendations

- **Targeted Mental Health Interventions:** Focus interventions on younger age groups and females.
- **Awareness Campaigns:** Educate social media users on the potential mental health impacts of prolonged usage.
- **Usage Guidelines:** Encourage balanced social media usage through screen-time limits.

X. Future Scope

This study lays the groundwork for exploring the relationship between social media engagement and mental health, but there are several opportunities to extend and deepen the analysis. Future research can leverage advanced techniques and expanded datasets to provide more actionable insights and address current limitations.

- **One potential direction is the application of Principal Component Analysis (PCA)** to reduce the dimensionality of the dataset and identify the core components driving mental health outcomes. By simplifying the dataset, PCA can help highlight the most critical variables influencing mental health, enabling more focused analysis and improving the interpretability of predictive models.
- **Predictive modeling** offers another avenue for future exploration. Machine learning models, such as logistic regression, random forests, or neural networks, can be developed to predict mental health outcomes based on social media usage patterns. Incorporating cross-validation techniques will ensure model reliability and robustness, while identifying high-risk individuals who may benefit from targeted interventions.
- **Conducting longitudinal studies** is essential for understanding how mental health metrics evolve over time. Collecting and analyzing temporal data can help establish causal relationships between social media usage and mental health outcomes, providing a dynamic perspective on the interplay between digital behaviors and well-being.
- **Developing interactive dashboards** is a practical application for stakeholders, enabling real-time visualization of the relationships between social media usage and mental health indicators. These dashboards could facilitate data-driven decision-making for mental health professionals, policymakers, and researchers, allowing for timely interventions and personalized recommendations.
- **Finally, expanded analysis** incorporating additional demographic factors, such as education level and occupation, can provide deeper insights into how social media affects different groups. Comparing the impact of specific platforms, such as Instagram versus TikTok, could reveal platform-specific patterns and risks, further enriching the understanding of social media's influence on mental health.

References

[1] M. Garg, "Mental Health Analysis in Social Media Posts: A Survey," *Advanced Intelligent Systems*, vol. 10, Jan. 2023. Available: https://link.springer.com/article/10.1007/s11831-022-09863-z

[2] M. De Choudhury, M. Gamon, S. Counts, and E. Horvitz, "Mental Health Analysis Via Social Media Data," in *IEEE Conference on Social Media*, 2013. Available: https://ieeexplore.ieee.org/document/8419435

[3] J. A. Naslund, A. Bondre, J. Torous, and K. A. Aschbrenner, "Social Media and Mental Health: Benefits, Risks, and Opportunities for Research and Practice," *Springer*, vol. 5, no. 2, Apr. 2020. Available: https://link.springer.com/article/10.1007/s41347-020-00134-x

[4] S. Galea and G. J. Buckley, "Social Media and Adolescent Mental Health: A Consensus Report of the National Academies of Sciences, Engineering, and Medicine," *PNAS Nexus*, vol. 3, no. 2, Feb. 2024. Available: https://academic.oup.com/pnasnexus/article/3/2/pgae037/7614669

[5] R. Safa, S. A. Edalatpanah, and A. Sorourkhah, "Predicting Mental Health Using Social Media: A Roadmap for Future Development," *arXiv*, Jan. 2023. Available: https://arxiv.org/abs/2301.10453

[6] S. S. P. Rao and K. S. R. Anjaneyulu, "Role of Exploratory Data Analysis in Data Science," in *IEEE Transactions on Data Science*, 2021. Available: https://ieeexplore.ieee.org/document/9488986

[7] J. Hullman and A. Gelman, "Exploratory Data Analysis," *Columbia University*, 2020. Available: https://stat.columbia.edu/~gelman/research/published/hullman_gelman_hdsr.pdf

[8] R. Bevans, "Hypothesis Testing: A Step-by-Step Guide with Easy Examples," *Scribbr*, 2019. Available: https://www.scribbr.com/statistics/hypothesis-testing

[9] SAGE Publications, "Introduction to Hypothesis Testing," *SAGE*, 2019. Available: https://us.sagepub.com/sites/default/files/upm-binaries/40007_Chapter8.pdf

[10] K. Singh and M. Xie, "Bootstrap: A Statistical Method," *Rutgers University*, 2008. Available: https://statweb.rutgers.edu/mxie/RCPapers/bootstrap.pdf

[11] M. Al Luhayb, F. Coolen, and T. Coolen-Maturi, "Smoothed Bootstrap Methods for Hypothesis Testing," *Springer*, vol. 8, 2023. Available: https://link.springer.com/article/10.1007/s42519-024-00370-x