Social Media and Mental Health: A Correlational and Predictive Analysis

Paul David Anghel[1], Andy el Asri[1], Ajay Baldewsing[1], and Hieu Viet Nguyen[1]

Vrije Universiteit Amsterdam, The Netherlands
Student Numbers: 2877964, 2655028, 2753103, 2760587
p.d.anghel@student.vu.nl, a.a.el.asri@student.vu.nl,
a.v.baldewsing@student.vu.nl, h.nguyen.viet@student.vu.nl

Abstract. Social media has become an important part of society and the lives of many people, raising concerns about its impact on mental well-being. This study examines the relationship between social media use and mental well-being and explores whether mental health can be predicted from digital habits. Using the Social Media and Mental Health Balance dataset from Kaggle.com [8], we combine correlational analysis with predictive modeling.
The results show strong correlations between social media use and mental well-being. Higher screen time is linked to higher stress levels and poorer sleep quality, which suggests that a more intensive social media engagement is associated with lower outcomes of mental well-being. Platform-specific analysis, further suggest that the mental-health impact of social media varies depending on the type of platform used.
In addition, the multi-output Ridge Regression model achieved moderate predictive accuracy, with stress being the most reliable predicted variable. Overall, the findings indicate that social media use can give meaningful insights with respect to mental well-being and highlight the importance of balanced digital engagement.

Keywords: Social Media · Mental Health · Predictive Modeling · Exploratory Data Analysis · Stress Level · Happiness Index · Sleep Quality.

1 Introduction

In recent decades, social media has become an important part of society and the lives of many people, especially young people. Social media refers to web and mobile platforms that allow individuals to connect with others within a virtual network where they can share, create, or exchange various forms of digital content [7].

As social media use continues to increase globally, concerns about the potential psychological consequences of its use are also growing. Many studies have already been conducted on the relationship between social media and mental well-being, but current debates are largely anecdotal. It is known that social media affects well-being, but there is a lack of predictive, data-driven insights [6]. Furthermore, numerous articles have shown that there are both negative and positive effects on mental health [5]. That's why we move beyond asking "What kind of impact is there?" to "How can we quantify and predict its impact?".

To fill this gap, this study addresses two research questions:

- **RQ1:** What is the correlation between social media use and mental well-being?
- **RQ2:** Can we build a model to predict mental health from digital habits?

In order to answer our research questions, we employ a twofold methodological approach. First, we statistically map the relationships between screen time, lifestyle, and mental health metrics. Second, we apply machine-learning regression models to forecast well-being. In doing so, we make use of the Social Media and Mental Health Balance dataset from Kaggle.com [8], which provides a rich foundation of behavioral and psychological variables. By combining correlational analysis with predictive modeling, this study offers actionable insights for billions of individuals that use social media to foster healthier and more informed digital habits.

The paper is organized as follows. Section 2 reviews the motivation. Section 3 describes related studies. Then section 4 describes the considered methodology while the results are presented in Section 5. We address the research questions, discuss recommendations and highlight our limitations in Section 6, and conclude in Section 7.

2 Motivation

Social media is part of almost everyone's routine, but we still don't have a clear picture of how it actually affects mental health well-being. A lot of the discussions surrounding mental health and social media happen on news sites, forums and social media like TikTok where people share personal experiences, but this is purely anecdotal and leaves out solid data science. Even in research, many studies focus on very broad variables such as screen time and then try to link them to broad terms like low mood. A problem with these studies as well is that these results jump around a lot too; sometimes heavier social media use is linked to worse mental well-being while sometimes it doesn't and it often depends on what people are doing online in the first place. So we're left with a lot of claims but not much clarity.

At the same time, the data people produce through everyday social media use has the potential to offer much clearer insight into mental well-being. Recent work shows that online behavior can reflect psychological states, but these predictive methods are almost never combined with traditional correlational research. These are highlighted in the Related work section. What is missing is a study that brings both approaches together: one that identifies which specific behaviors are related to well-being and also tests if those behaviors can be used to predict mental health outcomes. A mix of psychological and statistical research. This paper aims to fill that gap by using both statistical analysis and machine learning to better understand how social media habits relate to mental well-being.

3 Related Work

The relationship between social media usage and mental health has attracted much scholarly research in the past decade as platforms such as Facebook, Instagram and Twitter have become deeply connected to daily life. Early research focused on descriptive and cross-sectional analysis, but as the field progressed, it became clear that effects of social media on mental health are not uniform but depending strongly on how people use social media.

A large section of correlational research on this topic has linked heavy social media use with poorer mental health. A recent meta-analysis of student populations found that social media addiction was moderately associated with anxiety and depression and negatively associated with self-esteem [1]. A systematic review of adolescents and young adults reported that higher frequency use often corresponds with increased depression, anxiety and general psychological symptoms [2]. These effects were shown to be stronger for girls. But not all studies point in the same direction. A wide ranging systematic study showed that social media use also provides benefits such as the feeling of being connected to other persons [3]. These mixed outcomes suggest that the mental health impact of social media depends on how individuals use it.

More recent research have started to move beyond just correlational analysis to more complex, predictive, longitudinal analysis such as using machine learning to predict mental health. For example, one study investigated the use of neural user embeddings derived from Twitter post histories to quantify mental health status and the model was able to pick out users with depression and PTSD from control users [4]. This indicated that social media posts can have predictive signs of mental health into it.
Our study builds on both research fields, correlational research as well as predictive modeling, by looking at how different social media usage habits relate to mental health and by testing if these patterns can help predict mental well-being. By using the 'Social Media and Mental Health Balance' dataset, we combine correlational analysis with predictive modeling. This lets us see which online habits are linked to mental health while also checking if those habits can be used to predict mental well-being in a meaningful way.

4 Methodology

4.1 Subjects Selection

The Kaggle Social Media and Mental Health Balance dataset [8] we use for data analysis comes from a survey of 500 people between 16 and 49 years old. The group includes a mix of ages and backgrounds and their social media habits vary quite a bit. The survey asked persons about daily screen time, the social media platforms they used the most, their sleep quality, stress level, days they spend away from social media, how often they exercise and a happiness rating, all on a 1-10 scale. Most participants said they mainly use social media platforms like Instagram, Tiktok, Youtube, Facebook, X or Linkedin.

4.2 Experiment Design

The aim of our study is to see how social media habits of people relate to their mental well-being. Because the dataset includes measures of screen time, sleep, stress, exercise and happiness, we can look at both lifestyle factors as well as mental health indicators.
The analysis has two steps. First we run an exploratory data analysis to get an idea of the patterns in the data and to check with variables seem to relate to each other. After that, we build a predictive model to see if social media habits of new hypothetical users can be used to estimate their mental health scores based on their digital footprint. The technical work is done in a Jupyter notebook using Python and libraries such as Pandas, Numpy, Scikit-Learn and Matplotlib.

4.3 Data Analysis

The first part of the analysis looks at the data in a descriptive way. We compare habits like daily screen time and preferred platforms with mental well-being. With statistical analysis we check for patterns, like the correlation between screen time and stress, and visualize it using tools like box plots and heat maps. In the second part, we build a predictive machine learning model using Ridge Regression to estimate three mental health outcomes: sleep quality, happiness index and stress level. We prepare the data by scaling the numerical features (age, screen time, days offline, exercise frequency) and one-hot encoding the categorical features (gender, platform). The model is trained with an 80/20 train-test split, and it gets an average r^2 of 0.55, which means it can make moderately accurate predictions. We use a radar map and bar plots to show how different social media habits affect mental health when we train the model. Then, we are able to predict the mental health metrics of new user profiles that we create.

5 Results

5.1 Exploratory Data Analysis

5.1.1 Overview

This section presents an exploratory data analysis (EDA) of the Mental Health and Social Media Balance dataset, consisting of 500 individuals. The major purpose of this analysis is to find connections between digital behaviors, such how much time people spend on social media and the platforms they like best, and

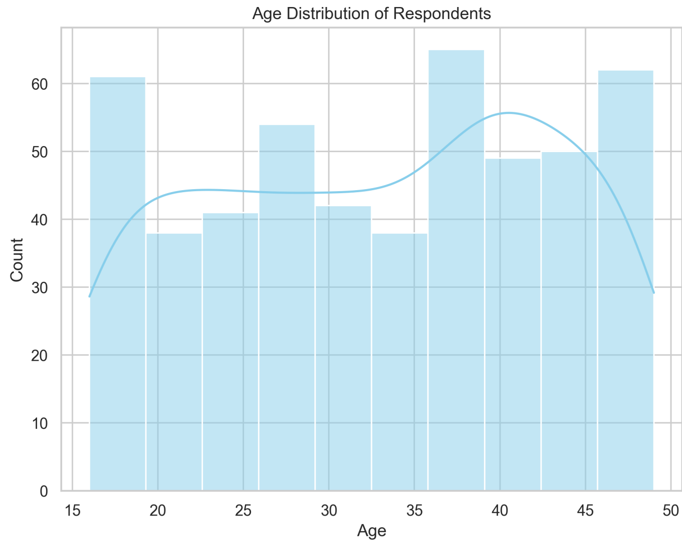


Fig. 1: Distributions of Age

significant mental health factors, like stress, happiness, and sleep quality. (Figure 1)

The dataset provides a diverse demographic , with respondent ages ranging from 16 to 49 years. The analysis was divided into three stages:

1. Analysis of Uni-variate Data: analyzing the distribution of individual factors, such as the amount of usage on particular platforms and the distribution of daily screen time, in order to evaluate the population's characteristics.
2. Bi-variate Analysis: looking at the direct links between how people use social media and how that affects their mental health, such comparing stress levels on different platforms.
3. Multivariate Analysis: using correlation matrices to identify complex links between lifestyle factors (such sleep and exercise) and mental health.

A preliminary examination indicates a group characterized by notably elevated self-reported happiness scores (Mean = 8.38/10), while simultaneously displaying considerable variability in stress levels. The study shows that the average daily screen time is about 5.53 hours, but the way people use social media and how it affects their mental health varies greatly depending on the site they use. The following subsections provide more information about these results, along with graphs and charts that support them.

5.1.2 Social Media Usage Habits

On average, respondents reported spending approximately 5.53 hours per day on social media platforms. To put this in context, if they are up for an average

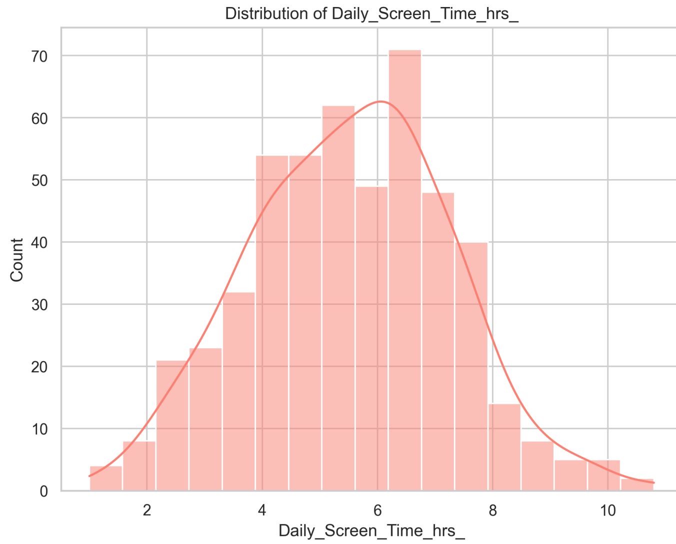


Fig. 2: Distributions of Screen Time

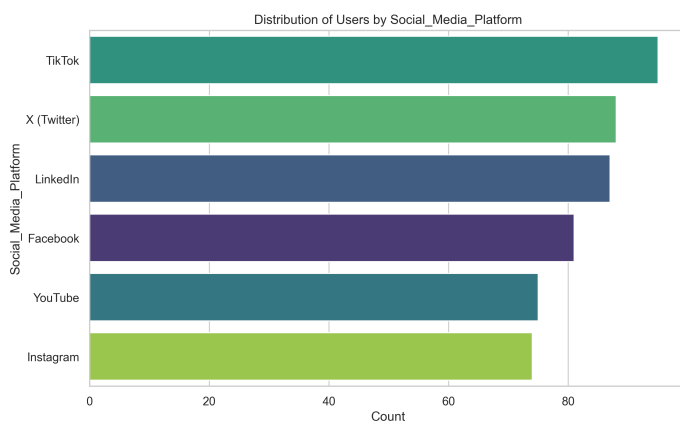


Fig. 3: Platform distribution

of 16 hours, this group spends about 35% of their time interacting with others online. The data also shows that there is a group of heavy users. The top 25% (the third quartile) said they used the service for more than 6.7 hours a day, and some even used it for almost 11 hours. This distribution indicates that for a substantial segment of the population, social media constitutes not only a recreational pastime but a primary daily engagement. (Figure 2)

TikTok was the most popular platform in our sample (around 19% of users), followed closely by X (Twitter) and LinkedIn. This order illustrates that people choose short, interesting content (like TikTok) and feeds with a lot of information and networking (like X and LinkedIn) above traditional social networking. Instagram, which is often called one of the top platforms in the world, was the least common in our sample. This departure from global trends may indicate unique demographic characteristics specific to this sample group or a shift in the preferences of the respondents. (Figure 3)

5.1.3 Mental Well-being Profile

Happiness:
The happiness levels stated in the sample are quite high, with a median score of 9 out of 10 and a mean score of 8.38 out of 10. This bias toward high happiness could mean that the sample group is particularly robust or that there is a bias in self-reporting, where people feel like they have to report higher levels of pleasure.

Stress & Anxiety:
Although the satisfaction scores were high, the stress levels were very different from person to person. Many people said they were not very stressed, yet another group of people said their stress levels were 8/10 or higher. This difference shows that the group is mostly happy, but certain things, such as their internet habits, make some subgroups more stressed out.

5.1.4 Screen Time vs. Mental Health

Our correlation heatmap reveals a distinct and concerning relationship between the duration of screen time and mental health metrics. (Figure 4)

- **Stress:** There is a strong **positive correlation (+0.74)** between Daily Screen Time and Stress Level. This implies a dose-dependent correlation: as people allocate more time to social media, their self-reported stress levels tend to rise considerably.
- **Sleep:** In contrast, there is a strong **negative correlation (-0.76)** between Daily Screen Time and Sleep Quality. This shows that people who use social media more often tend to have worse sleep hygiene and quality, which is often a sign of other mental health problems.

To further investigate these coefficients, we visualized the individual data points (Figure 5). These scatter plots confirm that the relationship is not merely statistical, but linear and consistent between genders.

- **Sleep "Drop-off":** As shown in Figure 5a, favorable sleep quality (scores of 7+) is observed among users with fewer than 4 hours of screen time. Beyond the 6-hour threshold, instances of high-quality sleep become almost non-existent, suggesting a limit to how much digital usage the sleep cycle can handle.
- **The Stress "Ceiling":** In contrast, Figure 5b illustrates a saturation of high stress levels among heavy users. Individuals who log more than 8 hours of daily screen time report almost exclusively stress levels in the 8–10 range.

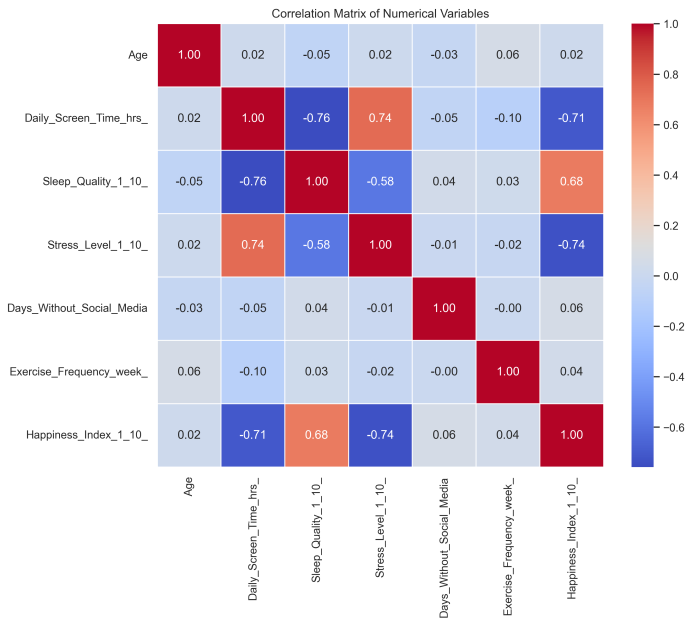
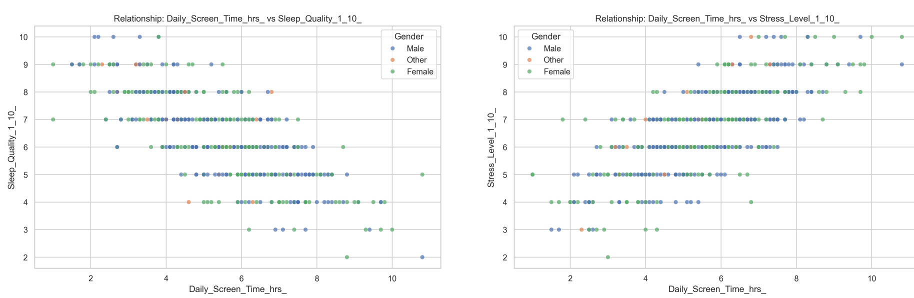


Fig. 4: Correlation matrix showing strong relationships between Usage, Stress, and Sleep.



(a) Impact on Sleep Quality (b) Impact on Stress Levels

Fig. 5: Scatter plots detailing the linear relationship between daily screen time and mental well-being indicators.

5.1.5 Platform-Specific Impacts

Seeing how stressed and happy people are on different platforms makes it clear that the user experience is very different on each one. Not all screen time affects users in the same way. (Figure 6)

- **Stress by Platform:** People who use Instagram and X (Twitter) said they were the most stressed out (Median ∼7-8), which could mean that these visu-

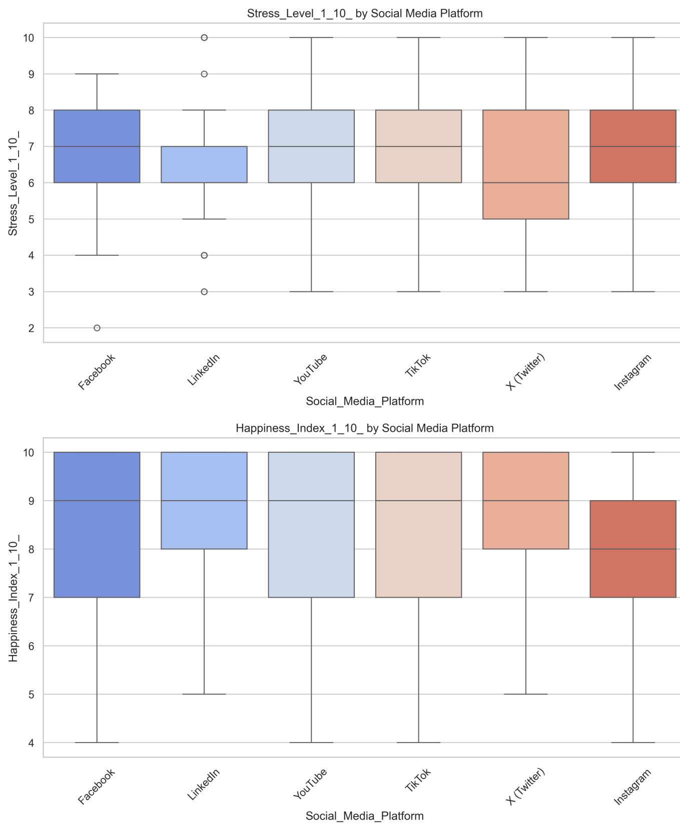


Fig. 6: Boxplots comparing Happiness Indices and Stress Levels across various platforms.

ally driven or text-heavy debate platforms are making people more stressed. Users of LinkedIn and YouTube, on the other hand, said their median stress levels were much lower.

- **Happiness by Platform:** Happiness scores also changed depending on the platform. The median happiness was generally high, but the distribution showed that users on Instagram had a wider range of scores. This means that their experiences were less stable and their emotional outcomes were less stable than those on more stable platforms like YouTube.

5.1.6 The Role of Lifestyle

The data shows that there is a very small or no link between Exercise Frequency and the mental health variables (Happiness, Stress) in this sample. This suggests that, for this group, the frequency of physical exercise was not a primary determinant (or alleviator) of their mental well-being scores, in contrast to the significant influence of screen time.

5.2 Predictive Modeling

5.3 Overview

This research used a multi-target regression framework, in order to predict key mental health metrics such as stress level, happiness index, and sleep quality. It relied on a mix of social media activity and lifestyle factors that were used to train the model. For example: age, gender, daily screen time, days without social media, exercise frequency per week, social media platform. The chosen predictive model for this task is Ridge Regression, combined with a MultiOutputRegressor, because it is reliable and fits well with continuous score predictions. The dataset was prior preprocessed and cleaned, missing values were replaced and further optimization were done to make it reliable and ready for the model training stage. The dataset was split into two parts, 80% for training and the remaining 20% for testing, in order to assess the accuracy of the model.

5.4 Model Performance

The model showed moderate predictive performance, with the following average performance scores: Mean R^2 Score = 0.55, Average RMSE = 1.02, Average MAE = 0.81, indicating an intermediate level of predictive performance across all target variables. These results highlight that the model provides useful predictive capabilities and identifies meaningful relationships between lifestyle choices and the use of social media with respect to Mental Health. The model's performance varied across the different mental health metrics, the Stress factor displayed the most reliable predictive performance. Furthermore, the Happiness Index displayed slightly less stability due to personal biases reflected in the initial dataset. The Sleep Quality index fell in the middle for predictive accuracy. This behavior is most likely due to some limitations, lack of data insights from the dataset, and users bias; these will be discussed further in the limitations section of the.

5.5 Predictions For New Users

To evaluate the ability of the model to generalize, we produced predictions for five hypothetical new users who differ in age, gender, preferred social media platforms, and lifestyle habits. These profiles, ranging from 18 to 60 years old, were constructed to reflect a spectrum of increasingly healthy digital behavior. Spanning from an 18-year-old woman with extremely high screen use and no physical activity to a 25-year-old moderately active man who uses TikTok, a 35-year-old physically active woman who uses Twitter, and a 45-year-old man on LinkedIn with well-balanced daily routines. And a 60-year-old woman who mainly uses Facebook, has very low screen time, spending most of the days offline.

In Figure 7 it is shown the radar chart comparing the predicted mental health metrics for the new users, against the dataset average. Additionally, Figure 8 illustrates the profiles of the new users and the predicted outcome and the impact on mental health.

The individualized prediction profiles depicted in Figure 8 demonstrate that due to differences in age, the influence of digital behaviors on mental health is significant. For instance, the 18-year-old with high screen time use, while having low offline time and no physical exercise at all, is predicted to experience high levels of stress, lower sleep quality, and less happiness than a 25-year-old, who is experiencing significantly better results due to healthier physical behaviors. Participants in their 30s and 40s displayed a more balanced predicted mental health profile. This outcome likely comes from a moderate social platforms usage and adopting healthier habits. The 60-year-old displayed the strongest overall mental health profile, characterized by low stress and the highest expected sleep quality. This pattern is consistent with her very limited screen time, substantial periods spent offline, and engagement in some physical exercise, all of which appear to be beneficial.

Therefore, the data suggest that younger individuals are more susceptible to severe negative consequences due to excessive digital use than older individuals, who have the option to mitigate these negative outcomes by managing their digital behaviors more responsibly.

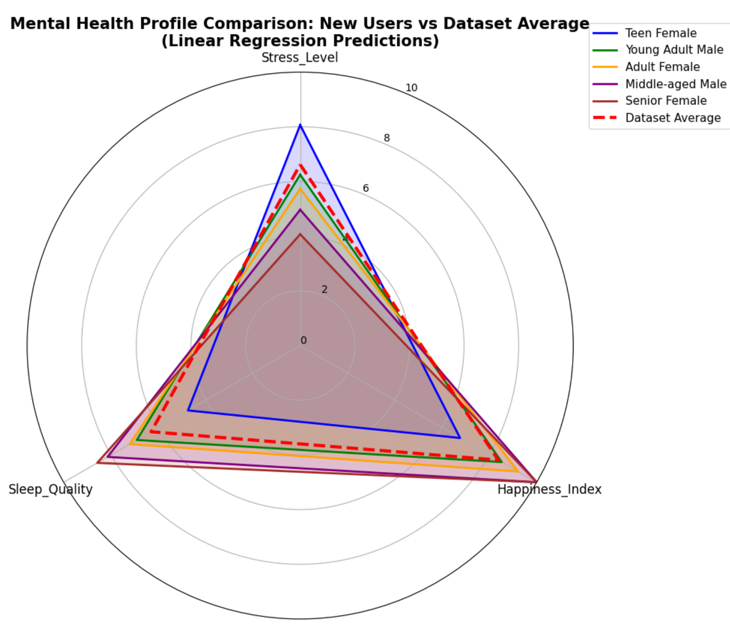


Fig. 7: Radar chart showcasing new user predictions with dataset average metrics

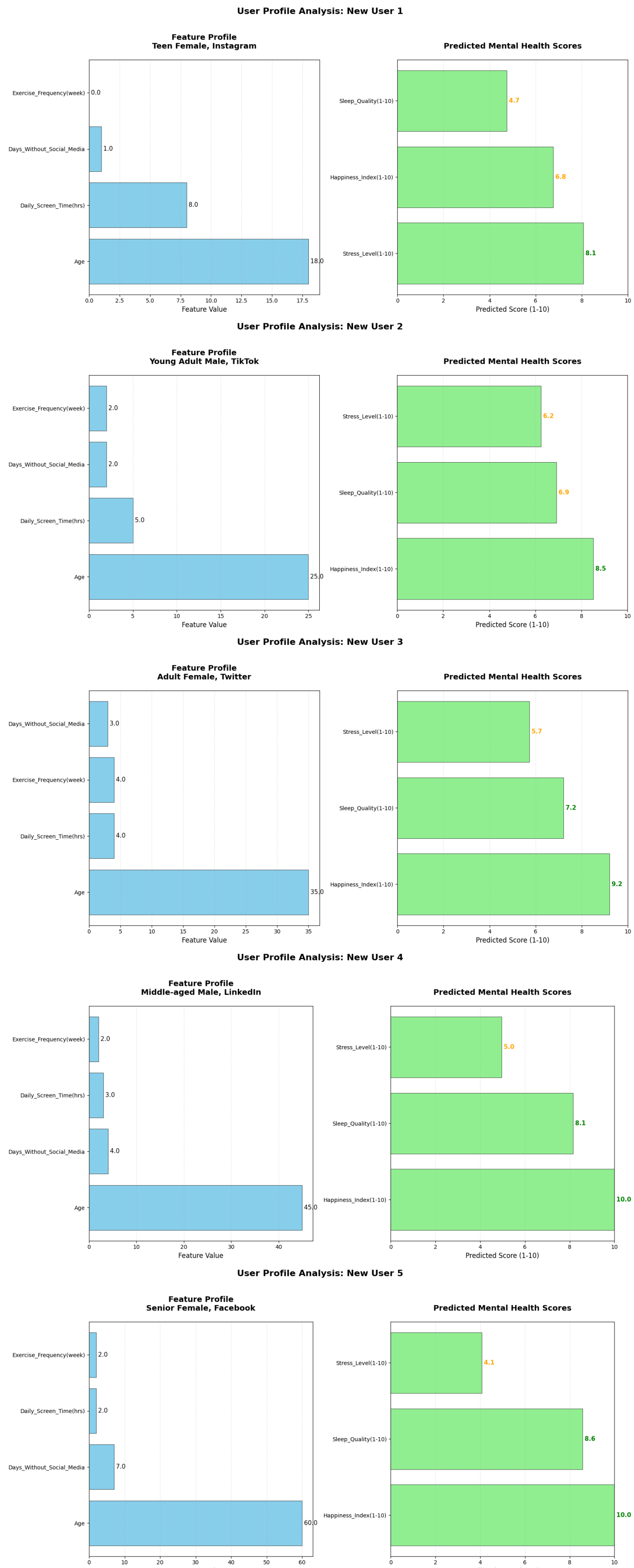


Fig. 8: Predicted mental–health outcomes for all five new user profiles.

6 Discussion

6.1 Addressing the Research Questions

This study sought to understand how social media habits relate to mental well-being. Our research focused on finding the correlation between social media use and mental well-being and if we could build a model to predict mental health from digital habits. Related to this, our study aimed to answer two central research questions:

RQ1: What is the correlation between social media use and mental well-being?
The results of the correlational data analysis show a clear and consistent relationship between social media use and mental-health outcomes. Daily screen time is strongly positively correlated with stress levels (+0.74) and strongly negatively correlated with sleep quality (-0.76). These findings indicate that heavier social media engagement is linked with higher stress and poorer sleep, which are both important factors regarding to mental well-being.

In addition, platform-specific analyses reveal that the impact on mental-health can differentiate for social media since users of Instagram and X reported higher median stress levels, whereas users of LinkedIn and YouTube showed more stable and generally lower stress scores. Therefore, these results suggest that also the context and type of social media plays an important role in influencing mental-health. Lifestyle variables such as exercise frequency showed little to no significant correlation with happiness or stress, which showcases that digital habits within this dataset played a more important role.

RQ2: Can we build a model to predict mental health from digital habits?
The results of the multi-output Ridge regression model was a moderate predictive performance, with an average R^2 score of 0.55. Stress was here the most reliable predicted variable. Furthermore, predictions for hypothetical new users demonstrated that the model can also differentiate mental health outcomes for different users who differ, among other factors, in age and preferred platform. Together, these results confirm that it is possible to create a predictive model, and digital habits are useful metrics to predict mental health factors such as sleep quality, happiness index and stress levels.

6.2 Recommendations

It's possible to give a couple recommendations based on our findings. One clear correlation is that people who spend a lot of time on social media show to report more stress and worse sleep. One of the recommendations for social media users would be to set boundaries around when and how long they use certain platforms. Not scrolling on social media right before heading to bed or in bed would help as well, as that is a prevalent practice among people. Moreover, including in their daily routine some physical activities and regular offline time will have a positive impact on the sleep quality, happiness and will reduce stress levels.

On the other hand, the predicted mental health profiles of new users should be taken objectively because it does not represent actual clinical trials and the individuals should consult their medical advisor in order to take informed decisions about their health and well-being.

6.3 Limitations

Even though the study gives a clearer picture of how social media habits connect to mental health, there are some limitations we should acknowledge. One of the limitations is that all the data comes from people filling in a survey about their own habits. This can introduce biases, since answering those questions depend on how honest a person is about his or her's social media usage. Some people might also guess or round numbers up and since factors like stress levels are subjective from person to person, rating scales might vary across users.

The sample size of 500 persons is also not fully representative for large populations, so it would be hard to extrapolate this study to big, diverse populations. People from other age groups or mixed backgrounds might have different online behaviors, so our findings do not extensively cover all possible users in large population sizes.

Furthermore, the predictive model has an average performance accuracy, presenting a limitation in accuracy potential that is not ideal. This outcome can be improved by using more powerful prediction algorithms such as Random Forest, Neural Networks, Gradient Boosting. Moreover, it would be better to have significantly more insights in the dataset regarding the digital metrics and the mental health factors in order to be able to cover a wider range of variations and correlations.

7 Conclusion

7.1 Main Findings

The main findings of this study indicate a strong correlation between social media use and mental well-being. Higher screen time is linked to higher stress levels and poorer sleep quality, which suggests that a more intensive social media engagement is associated with lower outcomes of mental well-being.

In addition, the predictive analysis demonstrated that mental-health outcomes can be estimated from digital habits with moderate accuracy with stress as the most reliable predicted variable. It is shown that the model is capable of differentiating mental-health outcomes across different patterns of users and digital behavior. The data suggests that having regular offline time and integrating physical exercises into the routine will have a positive impact on people's mental health. It seems that younger individuals tend to spend more time online than the older ones, and the middle aged group has more balanced habits due to being involved in other activities such as work, exercising.

7.2 Privacy and ethical consideration

This study may have used sensitive information related to mental health and digital behavior, which makes privacy and ethical consideration important. The data we used were publicly available and anonymized, so individual participants couldn't be identified. The predictive model developed in this research was intended solely for academic purposes and also given its moderate accuracy, it shouldn't be used for clinical diagnosis or advices regarding mental health.

7.3 Future work

Since the sample size of our research was not fully representative for large populations, future work could focus on collecting larger amounts of data from diverse backgrounds, so that this study would be representative of large populations. This could be expanded by collecting long-term data to better understand how relations between social media use and mental health influence each other over time.

Finally, future studies could explore more advanced machine-learning techniques for predictive models and investigate how predictive insights could develop healthier digital habits.

8 Individual Contribution

Andy el Asri: Conceptualization, Formal analysis, Investigation, Software, Writing
Ajay Baldewsing: Conceptualization, Formal analysis, Investigation, Software, Writing
Paul David Anghel: Conceptualization, Formal analysis, Investigation, Software, Writing (Results 5.2), created a Jupyter Notebook with Predictive Analysis findings
Hieu Viet Nguyen: Conceptualization, Data curation, Formal analysis, Software, Writing (Results 5.1), Jupyter Notebook with Exploratory Data Analysis (EDA) findings

References

1. Zhang, J., Wang, Y., Zhou, L., Wu, J., Li, H., Zhu, T.: Correlations between social media addiction and anxiety, depression, FoMO, loneliness and self-esteem among students: A systematic review and meta-analysis. https://doi.org/10.1371/journal.pone.0329466
2. Shannon, H., Bush, K., Villeneuve, P.J., Hellemans, K.G., Guimond, S.: Problematic Social Media Use in Adolescents and Young Adults: Systematic Review and Meta-analysis. https://doi.org/10.2196/33450
3. Taylor-Jackson, J., Moustafa, A.A.: The relationships between social media use and factors relating to depression. PMID: PMC7562923.

4. Amir, S., Coppersmith, G., Carvalho, P., Silva, M.J., Wallace, B.C.: Quantifying Mental Health from Social Media with Neural User Embeddings. https://doi.org/10.48550/arXiv.1705.00335
5. Sadagheyani, H. E., & Tatari, F. (2021). Investigating the role of social media on mental health. *Mental Health and Social Inclusion*, 25(1), 41–51.
6. Chancellor, S., & De Choudhury, M. (2020). Methods in predictive techniques for mental health status on social media: A critical review. *NPJ Digital Medicine*, 3(1), 43.
7. Naslund, J. A., Bondre, A., Torous, J., & Aschbrenner, K. A. (2020). Social media and mental health: Benefits, risks, and opportunities for research and practice. *Journal of Technology in Behavioral Science*, 5(3), 245–257.
8. Imran, A.: Social Media and Mental Health Balance. Kaggle Dataset. Available at: `https://www.kaggle.com/datasets/ayeshaimran123/social-media-and-mental-health-balance`