International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 11 | Nov - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Analysis of Social Media Impacts on Mental Health using machine learning

Ruvarashe Krystal Rusere [1], Kudzaishe Mazonde [2], Kuldeep Chouhan [3]

[1] *Department of Computer Science and Application, Sharda University*
[2] *Department of Computer Science and Application, Sharda University*
[3] *Department of Computer Science and Application, Sharda University*

Abstract - Technological advancement has made social media a central part of our lives, significantly affecting various parts of our mental well-being. In spite of it offering multiple opportunities for social interaction and providing unlimited learning resources, social media also has its share of advantages and disadvantages. Social media makes it possible to raise awareness about natural disasters, political, and social causes. The primary objective of this study is to analyse the relationship between social media usage and mental health using advanced machine learning techniques. This study made use of a publicly available Kaggle dataset with information on the social media usage and mental health data across 100 000 individuals. The study is expected to provide meaningful insights into the relationship between social media usage and mental well-being. The data set is analysed for a result that produces a better understanding of the correlation between users' digital habits and their mental health. The data set was cleaned, processed and analysed using Google Colab. After cleaning the data and processing the data, the Random Forest Regressor model was developed to predict mental health scores based on users' digital habits. The final results showed that the model achieved a Mean Absolute Error of 0.544, a Root Mean Squared Error of 0.809, and 0.596 as the R squared score achieving an averagely strong prediction accuracy. These findings provide a deeper insight of the correlation between social media usage patterns and mental well-being

Key Words: Dataset, Mean Absolute Error, Root Mean Squared, Data visualization, Random Forest

1.INTRODUCTION

In recent years, technological advancement has transformed the way people communicate, interact, and access information. At the center of this digital revolution lies social media, which has become an integral part of daily life for billions of individuals worldwide. Platforms such as Facebook, Instagram, Twitter (X), Snapchat, LinkedIn and TikTok provide unlimited opportunities for social interaction, networking, and knowledge sharing. For many, social media offers not only entertainment but also access to educational resources, professional development opportunities, and community engagement.

However, despite its widespread benefits, social media usage has also raised concerns regarding its potential negative influence on mental health. Studies have reported that individuals living with mental disorders or other severe mental illness, use social media platforms as comparable to other population with the use ranging from about 70% among middle age and older individuals [1]. The World Health Organization defines mental health as a state of well-being in which individuals recognize their abilities, handle everyday stresses effectively, engage in productive work, and contribute meaningfully to their community [2] Mental health is just as important as physical health and plays a vital role in overall well-being. The present study aims to analyze the relationship between social media usage and mental health through the application of advanced machine learning techniques. By leveraging publicly available datasets from Kaggle and other reliable sources, as well as collecting primary data through questionnaires administered to students, this research seeks to uncover patterns, trends, and sentiments in user behavior. Sentiment analysis will play a crucial role in interpreting textual content from posts and comments to better understand the emotional states reflected in online interactions.

To ensure clarity and accuracy, the collected data will undergo pre-processing and cleaning before being subjected to analysis. Machine learning models will then be employed to identify correlations and predictive patterns between social media activity and indicators of mental health. Finally, visualization tools such as Power BI will be used to present the results in a comprehensive and interpretable manner.

Ultimately, this research endeavors to provide valuable insights into the impact of social media on mental health, highlighting both its advantages and disadvantages. By combining data-driven methods with modern visualization techniques, the study will contribute to a deeper understanding of how social media usage influences psychological well-being and inform possible strategies for healthier digital engagement.

2. Methodology

For the processing of our research and analysis we used a step-by-step approach to effectively tackle each aspect. The following flowchart is a summarized representation of the steps we took.

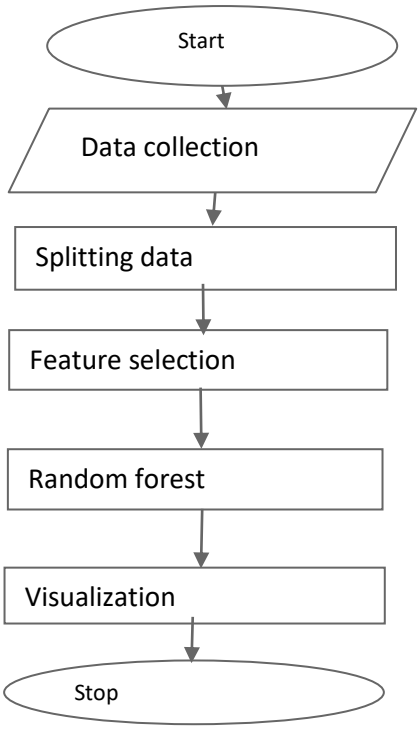

Fig 1

A. Description

This study employs a combination of quantitative analysis and machine learning techniques to explore the relationship between social media usage and mental health outcomes. A pre-existing dataset was incorporated in the analysis to help identify patterns and predict possible outcomes. The dataset in subject consists of questionnaire data primarily focused on social media habits and psychological state. For Better efficiency we used a pre-processed dataset. For the model development we particularly used a traditional ML model namely, Random Forest. In addition, the dataset was made use of under the appropriate data protection guidelines with no personal identifiers revealed. Data manipulation was handled through Pandas and numpy while matplotlib and seaborn were used for visualization

Random Forest algorithm

tonormalize = StandardScaler ()

X_train_scaled = tonormalize.fit_transform(X_train)

X_test_scaled = tonormalize. transform(X_test)

model = RandomForestRegressor (n_estimators=100, random_state=42)

model.fit (X_train_scaled, y_train)

y_pred = model. predict(X_test_scaled)

the_mae = mean_absolute_error (y_test, y_pred)

the_rmse = np. sqrt (mean_squarederror ((y_teerror (pred))

the_r2 = r2_score (y_test, y_pred)

3. Literature review

Fig 2

From the above figure is Okoli's guide approach which we implemented for our literature review. The steps below give detail to the above approach

1. Clearly identification of review purposes
2. Develop a detailed protocol document and provide training to the team to ensure consistency in the process
3. Practical screening to identify the studies to be included or excluded with justifications of any exclusions
4. Systematically extract literature research explaining how the search was conducted and ensuring comprehensiveness
5. Systematically extract relevant data
6. Establishingg criteria for excluding papers
7. Synthesizing the findings
8. Reporting the review

Studies show that excess use of social media leads to poor school performances []. Today we live in an intensively tech infused world in which people have grown to be dependent on social media. However, other researchers point out how important it is for people to not solely depend on results such as from ChatGPT until consulted by other verified sources.[4]. In other analyses, it has also been discovered that there is increase of anxiety and depression in cyber victims [5].

Year and authors			
	Objectives	***Methodology***	***Key findings***
2023 Vavek Bharwan	To understand the multifaceted effects of social media including performance,	Comprehensive analysis of existing	-Social media has multi-faceted effects -it offers

Year and authors	Objectives	Methodology	Key findings
	mental health and social interactions	research (exploring opinions and previous studies)	opportunities for learning and connection and entertainment but also causes distraction leading to poor academic outcome -there is need for a balance
2023 Arnkumar, so	to examine the effectiveness of social media as a tool and its impact on knowledge sharing	Quantitative research using surveys	-Social media has both positive and negative effects on students. -Positives include communication, networking -Negatives include false information, addiction and cyber bullying -social media promotes knowledge and increase student motivation
2021 Joshua Ebere	To discover The impact of social media platforms on student	Quantitative research using questionnaires for the determination of impact of social media platforms	-Social media platforms promote social interaction but also encourage physical distance among friends in meetings and events. -it brings limitations to the time spent with family and loved ones
2024 Collegenp	The impact of Social Media Student life	Comprehensive analysis of existing research) expert opinions and case studies	- Social media has complex and multi-faceted effects/impact on student life -Though it offers opportunities for learning, Connection & entertainment - It also causes distraction reduces study time and poor academic outcome. -There is therefore a need for a balance in terms of usage
Li et al., 2023	Detect cyberbullying driven mental distress	Graph neural networks	-Detection of distress is easier when social network interactions are included unlike when using text only
Haque et al., 2022	Estimating anxiety and depression severity	RoBERTa fine-tuning	-Domain specific pre training improves the scoring of symptom severity

Table 1

4. Results and discussion

A. Dataset

The dataset in subject examines the link between digital habits and mental well-being across 100000 individuals. It is inclusive of screen time tracking, social media usage (including TikTok) and the potential effects on the sleep quality, stress levels and mood. The dataset contains six pre-cleaned numerical features, well suited for machine learning applications such as regression and classification.[6]

B. Data Visualization

Below are the results of the data analysis conducted. Various charts, graphs, and heat maps were created to display the relationships between screen time, social media usage, and mental health outcomes, helping to interpret the findings in an accessible format that is also easy to understand.

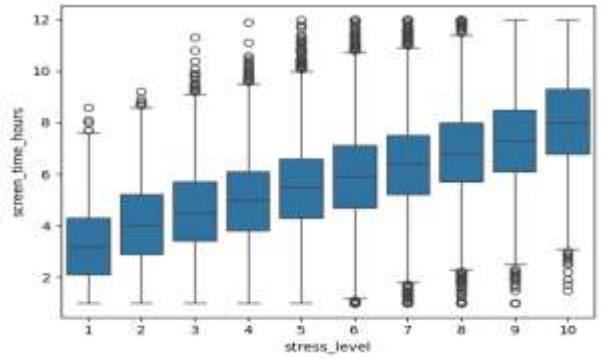

Fig 3

The box plot suggests that as the stress level increases the median time increases as well. However, the outliers indicated that some individuals are excessive users of their screens regardless of their stress levels. Generally, the low stress individuals usually have less screen time.

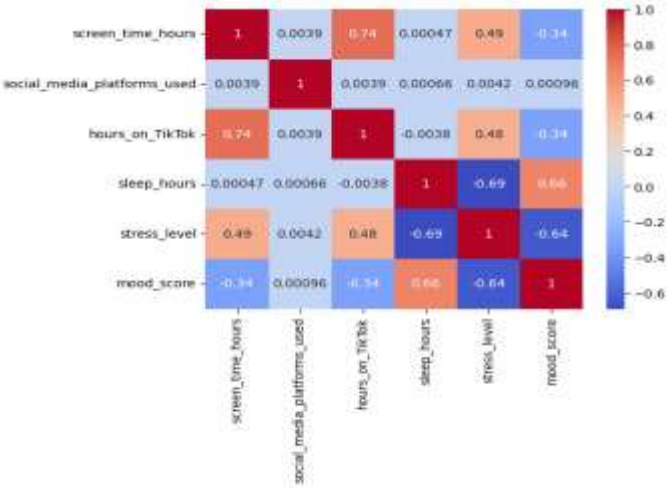

Fig 4

The above correlation map shows among other insights, the high positive correlation between stress level and tik-tok hours or screen time which indicates that high use of tik-tok is associated with increased stress levels minimal to no correlation with platform count is also observed from the results obtained.

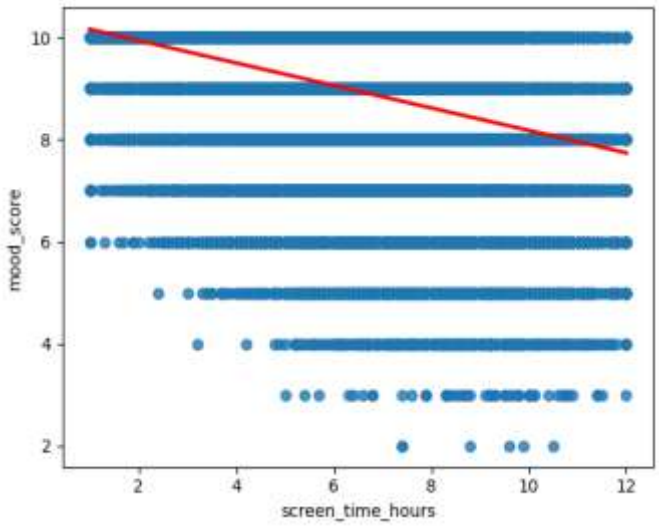

Fig 5

In figure 5, above, the red regression line slopes downward, indicating a negative correlation between screen time and mood score which in other words shows the inverse proportionality of the 2 factors. In addition, the lower mood scores clustered around high screen time proves the link between the excessive screen use and lower mood.

C. Data Modeling

The random forest regressor model was used in this study in order to predict the mental health scores and the model was trained with the data set Digital habits vs Mental Health data set.[6] In this study Random Forest Regressor was used because it works well with both numerical and categorical data [7] and it also works well with complex and non-linear relationships like social media usage, stress level and mood scores. Random forest works by combining multiple tress and calculating their average to find the final result. [4]. Using the 80:20 ratio to evaluate the performance of the model, the data set was divided into 80 % training and 20% for testing. The final result showed that the model Mean Absolute Error (MAE) of 0.544, a Root Mean Squared Error (RMSE) of 0.809, and an R^2 score of 0.596, indicating a moderately strong ability to predict mental health outcomes based on users' digital habits.

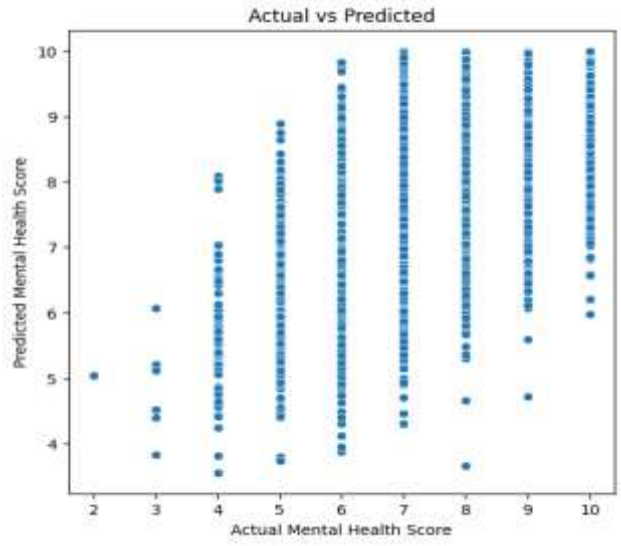

Fig 6

Fig 6 presents a scatter plot comparing the actual and predicted mental health scores generated by the Random Forest Regressor model. Each point in the graph represents an individual participant's actual mental health score plotted against the model's predicted score. The scatter plot indicates that as the actual mental health scores is directly proportional to the predicted scores which shows that the model successfully captured the overall relationship between social media usage and mental health.

However, the scatter along diagonal line suggests minor prediction errors, where the model either overestimated or underestimated certain values which overly aligns with the obtained performance results a Mean Absolute Error (MAE) of 0.544, a Root Mean Squared Error (RMSE) of 0.809, and an R^2 score of 0.596 which indicate that the model achieved a moderately strong prediction accuracy. Overall, the results demonstrate that the Random Forest Regressor is effective in identifying general patterns in the data, though there remains room for improvement in individual prediction precision

3. CONCLUSIONS

Based on the acquired results, we have seen that the use of social media plays a pivotal role in the psychological state of individuals therefore affecting mental health. Despite the good that social media contributes, we can conclude that excessive use of it proves be more problematic than helpful over time. Since we cannot shut down the use of social media as the use is inevitable in this age, we instead opt for a balanced and controlled use. Based on the model, which produced moderate results it indicates that not only social media is major contributor of mental health but there are other significant factors such as family environment, economical stress and academic pressure. The in cooperation of such a path will help manage a healthy mental state for individuals.

ACKNOWLEDGEMENT

We would like to appreciate the contributions that were offered by our guide, school and parents to make this paper a success. In addition, we would like to also appreciate the push from our faculties to achieve excellence in this research journey.

REFERENCES

[1] John A Nasland et al, "Social Media and Mental Health: Benefits, Risks, and Opportunities for Research and Practice" https://doi.org/10.1007/s41347-020-00134-x

[2] Ahmad Radwin et al," Predictive Analytics in Mental Health Leveraging LLM Embeddings and Machine Learning Models for Social Media Analysis https://dl.acm.org/toc/ijwsr-igi/2024/21/1

[3] Li, C., Managi, S. Mental health and natural land cover: a global analysis based on random forest with geographical consideration. Sci Rep 14, 2894 (2024). https://doi.org/10.1038/s41598-024-53279-7K. Elissa

[4] Li et al., 2023, "Detecting cyberbullying-driven mental distress using Graph Neural Networks"

[5] Haque et al., 2022, "Severity estimation of depression and anxiety from Reddit posts using RoBERTa"

[6] Jones Christydass, S.P., Nurhayati, N., & Kannadhasan, S. (Eds.). (2025). Hybrid and Advanced Technologies: Proceedings of the International Conference on Hybrid and Advanced Technologies (ICHAT 2024), April 26-28, 2024, Ongole, Andhra Pradesh, India (Volume 2) (1st ed.). CRC Press. https://doi.org/10.1201/9781003559139

[7] Jani, S.P., & Khan, M.A. (Eds.). (2025). Applications of AI in Smart Technologies and Manufacturing: Volume 2 (1st ed.). CRC Press. https://doi.org/10.1201/9781003682325