International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.102 ⨳ Peer-reviewed & Refereed journal ⨳ Vol. 14, Issue 4, April 2025
DOI: 10.17148/IJARCCE.2025.14401

The Impact of Social Media Usage on Mental Health: A Data-Driven Analysis

Manthan Narsingh Pandey [1], Touraj BaniRostam[2]

Student Master of Data Analytics, University of Niagara Falls, Niagara Falls Canada (UNFC), Niagara Falls, Ontario, Canada[1]

Assistant Professor, Master of Data Analytics, University of Niagara Falls Canada (UNFC), Niagara Falls, Ontario, Canada [2]

Abstract: **T**he introduction of social media has revolutionized how we communicate, interact socially, and conduct our daily lives worldwide. For as much as there are gains such as connectivity and access to information from social media, its possible negative contributions to mental health have been a concern. Several studies suggest that overuse of social media is associated with increased anxiety, depression, sleep issues, and poor concentration. Most of these claims have not been tested through critical examination of data, however. The purpose of this research is to investigate the link between social media usage and key mental health indicators through a quantitative, data-based methodology. This research includes both descriptive and prescriptive type of analysis which means that we have tried to offer some possible solutions based on the problems from the data. Our findings reveal significant associations between social media engagement and measures of mental health, with wide variation by gender. These results highlight the complex relationship that is present between virtual interaction and mental health. The results offer significant insights to policymakers, mental health professionals, and social media users, thereby indicating the necessity for responsible digital interaction and heightened mental health awareness. The findings of this study contribute to the knowledge of the multifaceted effect of online activity on psychological well-being. These findings have important implications for digital wellbeing policy, mental health practice, and responsible social media usage.

Keywords: Social Media, Mental Health, Data Analysis, Descriptive Analysis, Linear Regression

I. INTRODUCTION

In the modern age of digitization, social media has been woven into the fabric of daily life and influenced how individuals communicate, interact, and evaluate their self-worth. As much as it offers a wide range of advantages like real-time connectivity and access to vast information, its adverse psychological effects have drawn notable concern. Research has offered evidence of possible correlations between excessive use of social media and mental illness, anxiety, depression, sleep deprivation, and loss of concentration. Nevertheless, these relationships are still complicated and unclear, and it is hard to say to what degree social media influences well-being.

The growth in mental issues, particularly in youths, highlights the critical need to comprehend the effects of online behaviors on mental health. Overuse of social media can cause distraction, sleep deprivation, loss of self-worth because of comparison culture, as well as heightened anxiety levels. Yet, conventional research depends on self-reported questionnaires and qualitative accounts, which are susceptible to bias or lack a comprehensive data-centric perspective. With the support of data analytics, it may be conceivable to quantify these interactions, detect patterned behaviors, and provide evidence-based recommendations for more wholesome social media use [1].

Through analysis of actual user behavior and mental health data, it is possible to find patterns, identify warning signs at an early stage, and provide prescriptive recommendations that are personalized to individualized usage patterns. To investigate the complex interaction between social media use and mental health, we carried out an extensive data-driven study grounded on a structured dataset incorporating a variety of social media behaviors, psychological measures, and demographic measures[2].

To come up with meaningful insights, a stepwise analytical approach was employed that combined exploratory data analysis, correlation analysis, predictive modeling, and prescriptive analysis. Initially, Exploratory Data Analysis (EDA) was conducted by cleaning and preprocessing the dataset to impute missing values and categorical variables and then visualizing general usage patterns to identify significant trends [3].

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.102 ⪤ Peer-reviewed & Refereed journal ⪤ Vol. 14, Issue 4, April 2025
DOI: 10.17148/IJARCCE.2025.14401

We then performed a Correlation Analysis to ascertain the correlation between the duration of time spent on social media and the markers of mental health, including concentration issues, sleeping problems, and frequency of depression, to allow for the determination of potential correlations. Prescriptive Analysis was conducted in this analysis to provide evidence-based suggestions aimed at reducing distractions, enhancing sleep quality, and lowering anxiety, thereby providing actionable insights for healthier social media use.

Continuing, Section 2 provides a comprehensive review of relevant literature. In Section 3, we outline the Data Analysis process, which encompasses Data Preparation, Pre-Processing, and Expression. Section 4 focuses on Exploratory Data Analysis, where we detail our findings and achievements. In Section 5, we present the application of Predictive Modelling using Logistic Regression, highlighting the outcomes of our analysis. Finally, Section 6 offers a summary of the study and concludes with key insights and recommendations.

II. PAGE LAYOUT

Social media and Mental Health impacts have been studied extensively by researchers. According to past research, long-term use of social media is associated with high levels of stress and emotional instability, especially among young people. Whereas some studies highlight the negative psychological effects linked to long-term social media engagement, other research argues that its impact depends on usage habits, exposure to content, and personal variation. Despite these observations, a significant portion of existing studies rely on qualitative questionnaires and self-reported data, thus lacking empirical data-driven support.

Kailasam, V. K., et al., examine how social media can aid in suicide prevention. With suicide being a leading cause of death, especially among young people, 36% of victims leave a suicide note, often revealing feelings of apology or shame. Unfortunately, these notes are typically discovered too late. Social media platforms like Facebook, Twitter, and Tumblr, however, provide real-time opportunities for early intervention. The paper presents two cases where suicidal ideation was expressed on Facebook and proposes using social media for monitoring high-risk patients and improving suicide prevention strategies [4].

Fazida Karim et al., explore the impact of social media usage on mental health, focusing on anxiety and depression. After evaluating 16 papers selected from 50 shortlisted studies, the findings indicate that social media activity, including the time spent online, has a positive effect on mental health. However, the cross-sectional design and sampling limitations in most studies highlight significant variations in the results. The paper suggests that further research through qualitative methods and longitudinal studies is necessary to better understand the structure of social media's influence on mental health [5].

This study investigates the language used in Reddit posts about mental health, focusing on identifying linguistic characteristics that could be used to detect posts requiring urgent attention, such as those discussing suicidal intentions or harm to others. The findings reveal various discriminative linguistic features across mental health communities that could be useful in classification tasks. While negative sentiment is common across the dataset, the research also identifies condition-specific vocabularies used to communicate about different mental health disorders. These insights could help improve early intervention efforts on social media platforms [6].

This article explores the harmful effects of smartphones and social media on youth mental health. Studies show a link between heavy use and increased mental distress, self-harm, and suicidality, particularly among girls. Social media contributes to negative self-perception, cyberbullying, and exposure to harmful content, while also causing sleep deprivation and impairing cognitive and social functioning. The authors recommend clinicians work with families to reduce these harms and promote resilience through education and support. Public awareness and policy initiatives are also needed to create nurturing environments for youth [7].

Osman Ulvi et al, review the link between social media use and mental health, focusing on Facebook, Twitter, and Instagram. A meta-analysis of 39 studies, conducted from January 2010 to June 2020, reveals that while social media can foster a sense of community, excessive use, especially among vulnerable individuals, is linked to depression and other mental health issues [8].

Stevie Chancellor et al., examine how "Human-centered machine learning" (HCML) represents human subjects in mental health prediction on social media. Analyzing 55 papers, it identifies five key discourses—Disorder/Patient, Social Media,

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.102 ⁘ Peer-reviewed & Refereed journal ⁘ Vol. 14, Issue 4, April 2025
DOI: 10.17148/IJARCCE.2025.14401

Scientific, Data/Machine Learning, and Person—that create paradoxical portrayals of humans, potentially leading to dehumanization. The study highlights interdisciplinary challenges and provides guidelines for more ethical HCML research [9].

Ivan Sekulić et al, used deep learning methods to predict mental health status based on social media activity. By applying a hierarchical attention network for binary classification of nine mental disorders, the model outperforms previous benchmarks for four disorders. The study also analyzes word-level attention weights to identify key phrases relevant for classification and discusses the model's limitations [10].

This study explores the relationship between loneliness, identity styles, and internet addiction among students at the Faculty of Engineering, Islamic Azad University, Central Tehran Branch. In today's digital age, the internet plays a pivotal role in shaping people's lives, becoming an essential part of daily activities.

This descriptive survey study focuses on students from the Technical-Engineering faculties, using Multistage Cluster sampling. Three questionnaires were used to gather data on loneliness, identity styles, and internet addiction. Data analysis was conducted using SPSS software. The results revealed that students experience a higher level of loneliness compared to identity issues and internet addiction, suggesting that loneliness may be a contributing factor to these disorders [1].

III. DATA ANALYSIS

A. Data Preparation

The dataset integrates survey responses from online surveys with anonymized platform usage records. Participants provided self-reported demographic details, including gender, age, job, and marital status, along with an estimate of their daily social media usage. Usage logs covered multiple platforms such as YouTube, Instagram, Facebook, Twitter, TikTok, and LinkedIn. A preliminary data validation process ensured that column headers matched the expected data types, distinguishing between numerical and categorical variables. The primary dataset source can be accessed here.

B. Pre-Processing

Prior to analysis being conducted, a series of pre-processing steps were performed to maximize the quality and integrity of the data. Missing values, which can cause bias, were imputed based on the detection of missingness in demographic, social media usage, and mental health variables. Mean imputation for numeric data and mode imputation for categorical data were employed, and records with an unrealistically high count of missing values were deleted where required. Class imbalance, particularly for categorical variables such as gender, relationship status, and occupation, was addressed using several techniques, such as oversampling the minority class, under sampling the majority class, and using weighted analysis at the modeling phase. Noisy data, particularly for responses regarding social media platforms and mental health indicators, were addressed by eliminating inconsistent or erroneous entries, converting categorical responses to structured formats, and normalizing text-based entries for uniformity. Boxplots were used to detect outliers as a way to find extreme values in numeric columns such as "avg_time_per_day," while Z-score and interquartile range (IQR) methods helped in removing extreme deviations from the mean. In order to enhance the predictive power of the dataset, feature engineering was carried out by adding new features such as "Time Spent Per Platform" and "Social Media Influence Score." Furthermore, categorical variables were encoded with One-Hot Encoding or Label Encoding, and numeric values were normalized for consistency in regression analysis.

IV. EXPLORATORY DATA ANALYSIS

Exploratory Data Analysis (EDA) provided insightful information regarding the distribution and pattern of the dataset. The age distribution revealed that most respondents fell within the age bracket of 18 to 35 years, with a histogram revealing a slight right skew, suggesting that the dataset predominantly comprises youths. In terms of gender balance, the results showed a relatively balanced composition, with a marginally larger percentage of females than males, and a minor percentage of individuals who identified as non-binary. In terms of profession, a majority of participants were young professionals or university students, and fewer individuals were self-employed or represented non-traditional professions. The relationship status distribution showed that most participants were in a relationship or single, with fewer people indicating divorced or married. This aspect was also examined for potential correlations with social media behavioral patterns. Besides, the use of social media was determined by an analysis to entail the majority of users allocating two to four hours per day on such networks, with a considerable percentage having spent over five hours a day, thereby marking a high extent of engagement with social media.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁙ Peer-reviewed & Refereed journal ⁙ Vol. 14, Issue 4, April 2025

DOI: 10.17148/IJARCCE.2025.14401

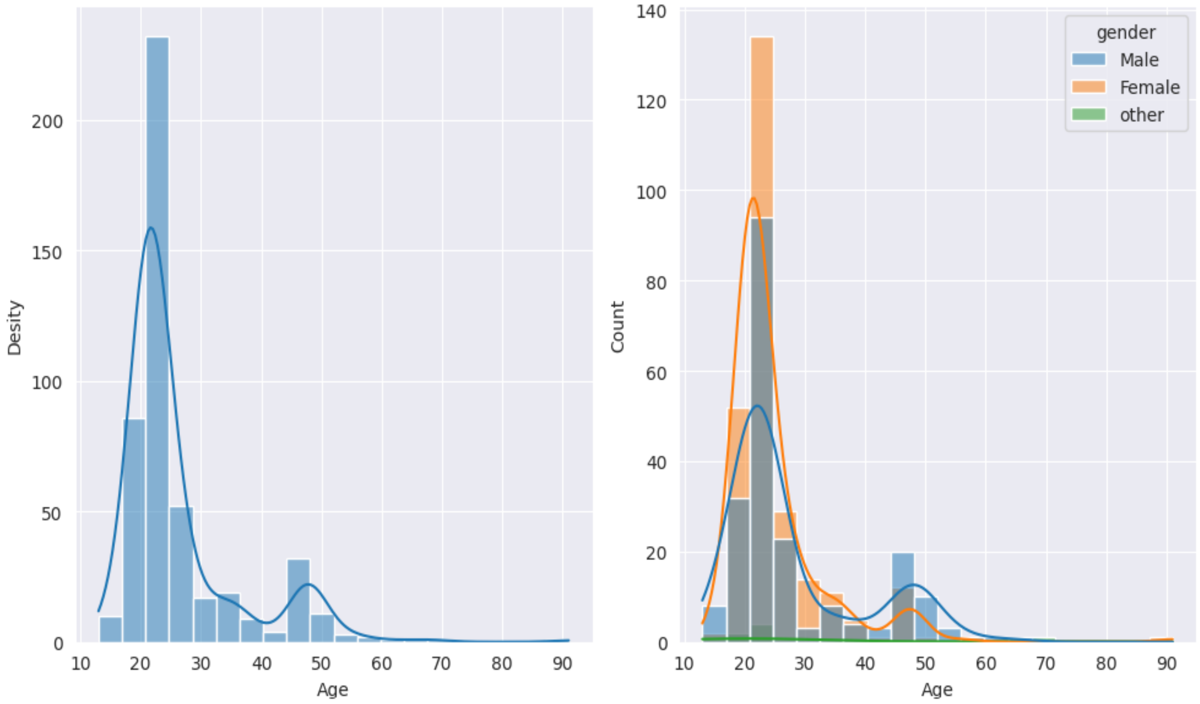

Fig.1 Age Distribution Graph

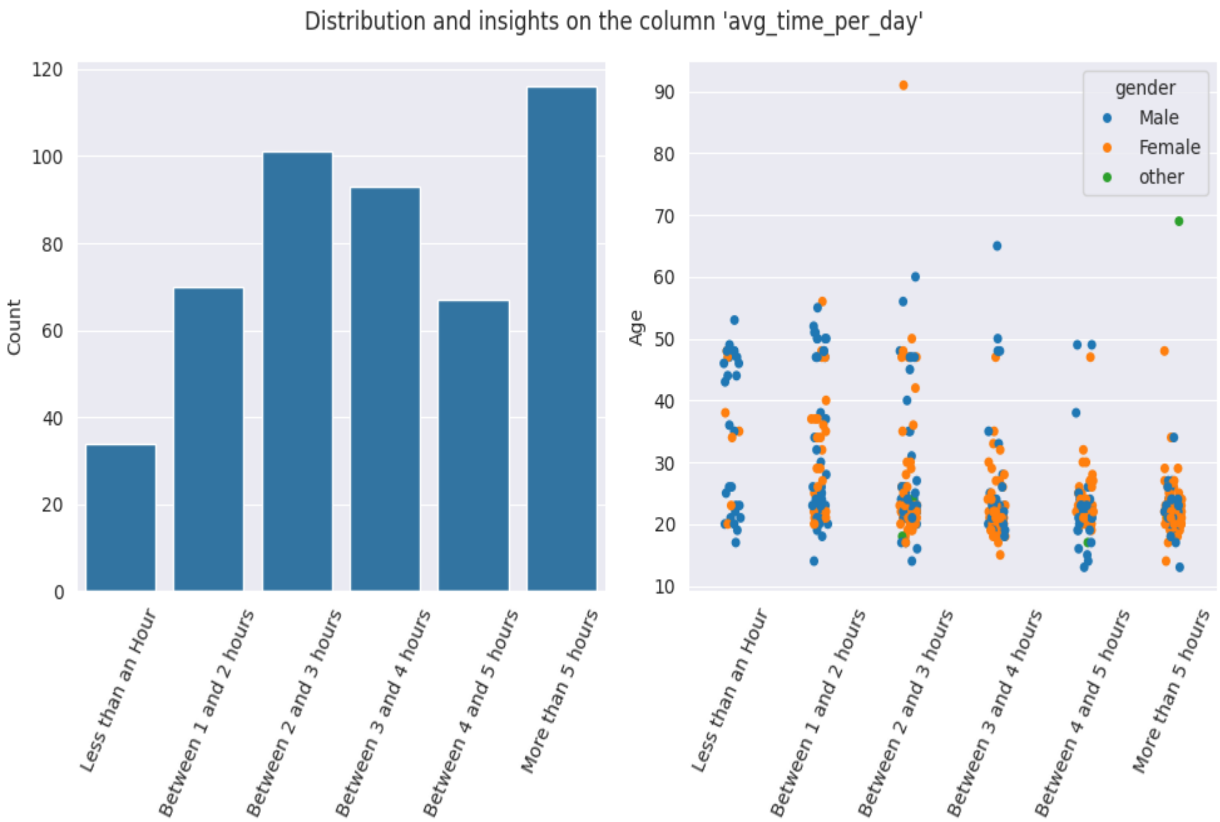

Fig 2 Distribution of Avg Time Per Day

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁘ Peer-reviewed & Refereed journal ⁘ Vol. 14, Issue 4, April 2025

DOI: 10.17148/IJARCCE.2025.14401

A platform-specific in-depth analysis gave insights into behavior that were unparalleled. Instagram, YouTube, and Twitter were the most used social media platforms, with declining Facebook use most evident among young people. On the other hand, platforms such as Discord and Reddit had a small but highly active user base.

When examining the platform usage by age group, it was clear that younger users between 18 and 25 years old mostly favored Instagram, TikTok, and YouTube, while older users over 25 years old showed higher usage of Facebook and LinkedIn. Interestingly, Twitter usage was constant across all age groups. When it came to time spent on social media activities, peak hours were found to be between 8 PM and 11 PM, with university students showing higher screen time compared to professionals. Moreover, a high percentage of users reported late-night social media use, a trend that was linked to sleep disturbance.

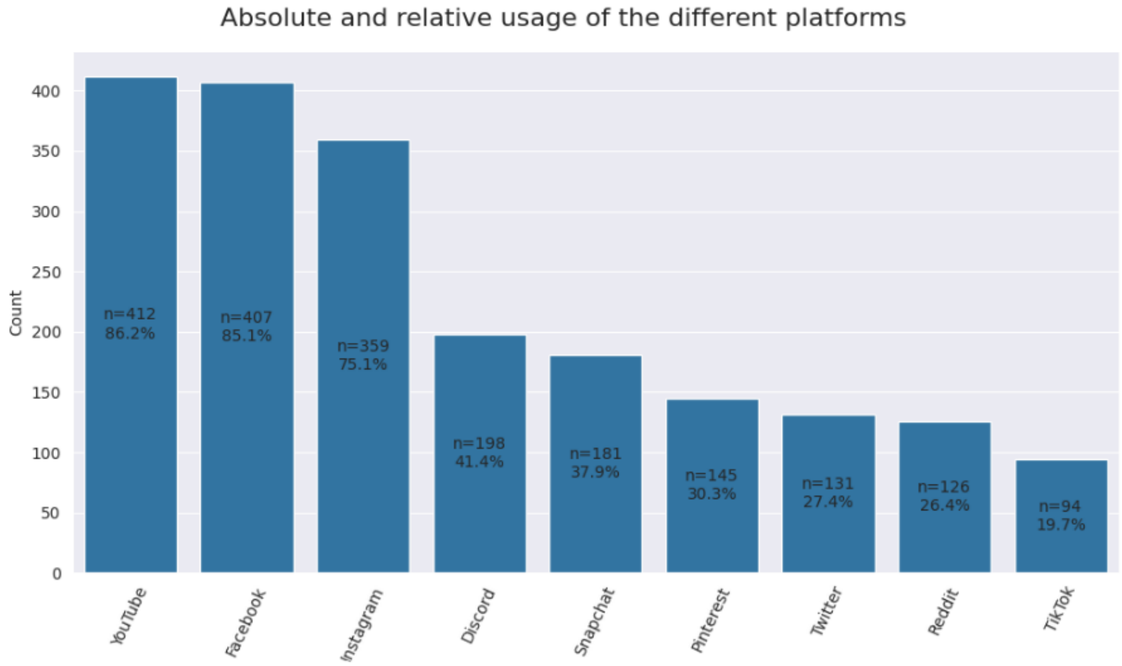

Fig 3 Usage of Platforms

V. PREDECTIVE MODELLING USING LOGISTIC REGRESSION

For this study, a logistic regression model was developed to predict the probability of negative mental health outcomes based on demographic factors and social media usage patterns. The data for this analysis included self-reported survey data, in which respondents indicated their age, sex, marital status, average daily hours spent on social media, and which specific platforms they use most frequently.

The dependent variable of risk was derived from an impact_sum score of 12-60, with the lower scores (12-36) being lower risk (0) and the higher scores (37-60) being higher risk (1).

Independent variables of the model included demographic factors of gender, age, and relationship status and social media data of average daily usage and activity on Pinterest, YouTube, Reddit, TikTok, Instagram, Facebook, Snapchat, and Discord. Preprocessing of the data, which involved replacing NaN values with zeroes before handling missing values, as well as categorical encoding to make categorical variables suitable for machine learning, was conducted before training the model.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.102 ⁙ Peer-reviewed & Refereed journal ⁙ Vol. 14, Issue 4, April 2025
DOI: 10.17148/IJARCCE.2025.14401

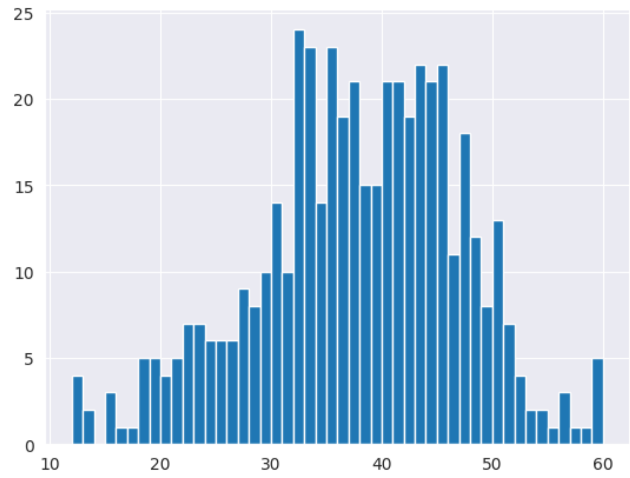

Fig 6.1 Impact Sum Class for Linear Regression

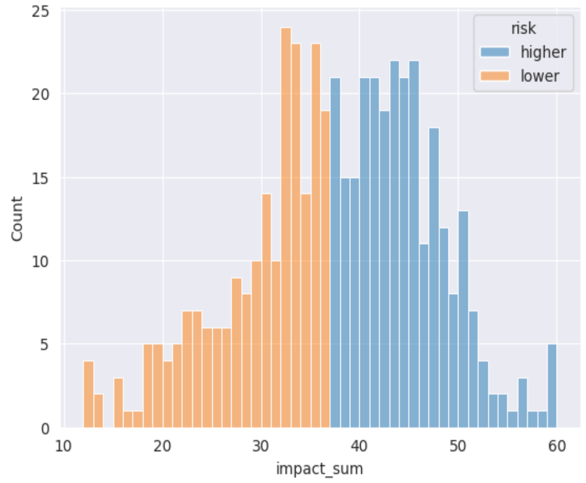

Fig 6.2 Division of Impact Sum Class

Specifically, the target feature, or "risk," was label-encoded into binary values, whereas categorical ones, i.e., "gender," "relationship," and "avg_time_per_day," were encoded based on the ordinal encoding approach. Then the data set was divided 80-20 for training data and test data with a guarantee of preserving class balance using stratified sampling.

Logistic regression was used as it has proven to be efficient in such binary classification problems, and this was trained using the lib-linear solver. Performance of the model was assessed using measures of accuracy and a confusion matrix to quantify classification quality. Results showed that the model was able to classify 70% accurately whether an individual is at risk or less likely to experience poor mental health outcomes from their use of social media and demographic circumstances.

This finding reflects that while social media usage patterns and demographics provide a moderate level of predictive power, they do not fully explain mental health risk, implicating the involvement of other psychological, social, or behavioral mechanisms.

The analysis of the confusion matrix also unveiled instances of misclassification, hence showing potential areas of improvement via feature engineering, incorporation of sentiment analysis of social media posts, or the application of more sophisticated machine learning algorithms such as Random Forest or Neural Networks.

The results of this research have significant implications for the following insights:

(1) certain social media websites may be more strongly related to mental health outcomes, which could be further examined through feature importance analysis;
(2) the amount of time spent on social media daily may be a significant predictor of mental well-being, thereby corroborating existing research connecting excessive screen time with mental health problems;
(3) demographic factors such as age, gender, and relationship status may moderate how individuals use social media and how it relates to their psychological well-being;
(4) depressive symptoms and sleeping problems were moderately correlated with increased screen time, possibly linking social media with certain mental health issues; and
(5) validation-seeking behavior was a strong indicator of high social media usage.

In sum, this research demonstrates that social media usage patterns, when combined with demographic characteristics, can be used to predict mental health risks with a high level of accuracy.

However, further research is needed to refine model performance through the inclusion of additional variables, including measures of psychological wellness, emotional states, and behavioral patterns, which can provide deeper insights into the relationship between social media and mental health.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 4, April 2025
DOI: 10.17148/IJARCCE.2025.14401

VI. CONCLUSION

This study sought to predict the risk of adverse mental health consequences from demographical characteristics and social media use patterns with a logistic regression model. By splitting individuals into high-risk and low-risk categories, the model had an accuracy rate of approximately 70%, meaning that although demographical characteristics and social media behavior have moderate predictive abilities, they are not the only determinants of mental health risk.

The results show that several factors, such as age, gender, marital status, daily screen usage, and use patterns of platforms, have important effects on mental health outcomes, with increased usage of social media possibly elevating risk levels. However, misclassification errors highlighted in the model highlight the importance of including extra variables, e.g., markers of psychological health, emotional status, and content-based sentiment measures, to increase predictive accuracy.

The results corroborate existing literature linking social media usage with mental health conditions and reinforce the complex, multifaceted nature of these relationships. Research in the future needs to examine more advanced machine learning techniques, longitudinal data, and qualitative results to more clearly determine how social media usage affects mental health.

Moreover, policymakers, mental health professionals, and educators could use these findings to craft targeted interventions, digital wellness programs, and awareness campaigns that would seek to encourage healthier social media usage. Ultimately, although social media occupies a prominent position in contemporary communication, its psychological impact needs ongoing investigation to ascertain that digital platforms catalyze healthy mental health trends as opposed to propagating risks.

REFERENCES

[1]. T. Banirostam, and H. Saberi, “Study of Loneliness and Identity Styles among Internet Addicted Students in Faculty of Engineering, Islamic Azad University, Central Tehran Branch”, *Emerging Science Journal*, vol. 1, no. 2, pp. 119, 2017, DOI: 10.28991/esj-2017-01119.

[2]. S. Mohammadi, and T. Banirostam, “A Perceptual Meta-Model Based on the Ontology of Mental Models”, *The International Journal of Humanities and Social Studies*, vol. 3, no. 11, pp. 122–128, 2015.

[3]. F. Yousefian, T. Banirostam, and A. Azarkeivan, “Prediction of Thalassemia Based on Artificial Intelligence Techniques: A Survey”, *Int. J. of Advanced Research in Computer and Communication Engineering*, vol. 6, no. 8, pp. 281-287, 2017, DOI: 10.17148/IJARCCE.2017.6847.

[4]. V. K. Kailasam, "Can social media help mental health practitioners prevent suicides?", *Current Psychiatry*, vol. 14, no. 2, pp. 37-39, 2015.

[5]. F. Karim, A. A. Oyewande, L. F. Abdalla, R. C. Ehsanullah, and S. Khan, “Social Media Use and Its Connection to Mental Health: A Systematic Review,” *Cureus*, vol. 12, no. 6, e8627, Jun. 2020, DOI: 10.7759/cureus.8627.

[6]. G. Gkotsis, *et Al.*, “The language of mental health problems in social media,” *Proceedings of the 3rd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, San Diego, California, Jun. 16, 2016, pp. 63–73, Association for Computational Linguistics.

[7]. E. Abi-Jaoude, *et Al.*, “Smartphones, social media use and youth mental health,” *CMAJ*, vol. 192, no. 6, pp. E136-E141, Feb. 2020, DOI: 10.1503/cmaj.190434.

[8]. O. Ulvi, *et Al.*, “Social Media Use and Mental Health: A Global Analysis,” *Epidemiologia*, vol. 3, no. 1, pp. 11-25, Jan. 2022, DOI: 10.3390/epidemiologia3010002.

[9]. S. Chancellor, *et Al.*, “Who is the 'Human' in Human-Centered Machine Learning: The Case of Predicting Mental Health from Social Media,” *Proceedings of the ACM on Human-Computer Interaction*, vol. 3, no. CSCW, Article No. 147, pp. 1-32, Nov. 2019, DOI: 10.1145/3359249.

[10]. I. Sekulić and M. Strube, “Adapting Deep Learning Methods for Mental Health Prediction on Social Media,” *Proceedings of the 5th Workshop on Noisy User-generated Text*, pp. 322-327, 2019, DOI: 10.18653/v1/D19-5542; arXiv:2003.07634, https://doi.org/10.48550/arXiv.2003.07634