International Research Journal of Multidisciplinary Scope (IRJMS), 2024; 5(2): 623-635

Original Article | ISSN (O): 2582-631X DOI: 10.47857/irjms.2024.v05i02.0592

Machine Learning Analysis of Social Media's Impact on Mental Health of Indian Youth

Vengalarao Pachava[1]*, Olusiji Adebola Lasekan[2], Siva Krishna Golla[3], Sreeramulu Gosikonda[1]

[1]NMIMS University, Hyderabad, India. [2]Departamento de Educación e Innovación, Universidad Católica de Temuco, Temuco 4780000, Chile. [3]School of Management Studies, National Forensic Sciences University, Gandhinagar 382007, India. *Corresponding Author's Email: pachavaedu@gmail.com

Abstract

This study delves into the complex relationship between various mental health indicators and their influencing factors among Indian youths. It specifically examines how external validation, interactions on social media platforms, and demographic variables such as gender, age, and occupation impact a range of mental health outcomes. These outcomes include experiencing negative thoughts, a disinterest in activities, low self-esteem, the development of eating disorders, disturbances in sleep patterns, symptoms of depression, difficulties in concentration, and feelings of fatigue. Employing the theoretical framework of Social Cognitive Theory, this research utilized an online random sampling method to gather data from a diverse group of 151 Indian youth participants. The findings of this study highlight the significant role that external validation and social media usage play in shaping mental health conditions among the youth. This underscores the critical need for integrating digital literacy components into mental health initiatives, aiming to foster healthier online behaviors and interactions. Furthermore, the study advocates for a holistic approach to mental health care, emphasizing the consideration of the specific needs of various demographic segments. It suggests that future mental health policies and interventions should be culturally sensitive and responsive. The results of this research underscore the pressing necessity for ongoing investigations into these vital dynamics, aiming to better understand and address the mental health challenges faced by Indian youths.

Keywords: Demographics, Machine Learning, Mental Health Outcomes, Indian Youths, Social Media Behaviour.

Introduction

There has been an increasing awareness in mental health research of the significance of different social and behavioral indicators in comprehending mental well-being. This is especially important in today's fast-paced world where societal norms are constantly evolving and social media has a strong impact. It is therefore imperative to examines mental health indicators across various factors such as gender, age group, occupation, external validation, and social media posting. It emphasizes the intricate relationship between these factors and their influence on mental health, particularly among young people in India. Extensive research has documented the existence of mental health disparities between genders, highlighting notable variations in the occurrence and nature of mental health problems among men and women (1, 2). For instance, women are more likely to suffer from anxiety and depression, while men may exhibit higher rates of substance abuse (3). These disparities underscore the need for gender-sensitive mental health interventions and policies. Age is a significant factor in mental health, as different stages of life offer their own set of challenges and stressors. During adolescence and young adulthood, individuals face important challenges such as developing their identity, building relationships with peers, and dealing with the pressures of academics or work (4). For the elderly, issues of loneliness, loss, and cognitive decline are more prevalent, impacting their mental well-being (5). Occupational stress greatly impacts mental health. Various occupations can be demanding and place people under immense pressure, which can result in stress, burnout, and mental health concerns. Healthcare workers and those in high-stress jobs are especially susceptible to these challenges (6, 7). The COVID-19 pandemic has further exacerbated these issues, highlighting the importance of workplace mental health support.

The impact of external validation, especially

(Received 09th February 2024; Accepted 25th April 2024; Published 30th April 2024)

through social media, has become a significant concern for mental well-being. Seeking validation and worrying about how others perceive us on social media platforms such as Instagram and Facebook can contribute to heightened anxiety, depression, and a skewed sense of self, particularly among teenagers and young adults (8). While these factors are increasingly acknowledged, there is a notable lack of research on how they collectively affect the mental health of young people in India. This population encounters distinct cultural, social, and economic pressures that can impact the expression and encounter of mental health concerns. There is a significant demand for in-depth research that analyzes these indicators in the Indian context, considering the country's diverse socio-cultural landscape and rapidly evolving digital environment (9, 10).

Therefore, this paper aims to investigate the effect of gender, age group, occupation, external validation, and social media posting behaviors on several mental health issues among young people in India. By exploring these relationships, the study seeks to contribute valuable insights into the multifaceted nature of mental health determinants in the Indian context, addressing a critical research gap in the current literature.

Using Social Cognitive Theory (SCT) as a theoretical framework provides a detailed perspective to examine mental health indicators across different demographics and behaviors. The significance of observational learning, self-efficacy, and triadic reciprocality in comprehending the formation of behaviors and attitudes towards mental health through gender, age group, occupation, external validation, and social media posting behaviors is emphasized in SCT (11). Observational learning suggests that individuals can adopt mental health-related behaviors by observing others, especially on social media platforms, impacting their mental well-being differently across gender and age groups (12). Self-efficacy, which is impacted by seeking validation on social media, has a significant impact on how individuals perceive their ability to handle stressors. This, in turn, affects their mental well-being and can differ greatly depending on their occupation (13). The concept of triadic reciprocality emphasizes the ongoing interplay between personal factors, behavior, and environmental influences, highlighting the complex interplay between individual characteristics, social media engagement, and occupational stressors in shaping mental health outcomes (11). This comprehensive approach underscores the need for targeted interventions that address the specific needs of different demographic groups in the digital age.

Observational learning, as rooted in Bandura's Social Cognitive Theory, explores how individuals acquire and imitate behaviors they see on social media, which can have a significant impact on their mental health attitudes and actions (11). This intertwines with self-efficacy, where individuals' belief in their capability to control their functioning and events in their lives is shaped by gender, age group, occupation, external validation, and social media posting behaviors. The concept of triadic reciprocality further elucidates the dynamic interaction among personal factors (such as gender, age, occupation), behavior (notably, social media posting), and environmental factors (like external validation), showcasing the mutual influence these spheres exert on one another (11). These personal factors, including gender, age group, and occupation, diversify experiences on social media, affecting mental health outcomes across different demographics (13). External validation, derived from social media engagements such as likes, comments, and shares, emerges as a critical environmental factor that affects self-efficacy and, by extension, mental health outcomes, highlighting the significance of social approval in the digital age (13). Social media posting behavior, influenced by and contributing to observational learning and self-efficacy, plays a crucial role in shaping mental health outcomes, with the nature of engagement and content consumed being pivotal (12). The interaction between observational learning, self-efficacy, triadic reciprocality, and personal and environmental factors leads to different mental health outcomes such as anxiety, depression, and stress levels. This highlights the need for a comprehensive approach in researching and implementing interventions to improve mental health in different populations (12). This comprehensive framework underscores the intricate relationship between social media use and mental health, highlighting the imperative for nuanced research and interventions to address mental health improvements across varying populations.

In recent years, the mental health of youths has emerged as a significant concern, underscored by a complex interplay of factors leading to a spectrum of mental health indicators, such as negative life thoughts, low interest in activities, low self-esteem, poor appetite or overeating, trouble sleeping, feeling depressed, trouble concentrating, and constant fatigue. This literature review embarks on an exploration of these indicators across various dimensions including gender, age group, occupation, external validation, and social media posting, aiming to provide a comprehensive snapshot of the prevailing knowledge landscape. Research has revealed significant gender disparities in the occurrence and manifestation of mental health indicators among youths. Women tend to experience a greater occurrence of depression and anxiety symptoms, including difficulties with sleep and low self-esteem, in comparison to men (14). These differences are often attributed to societal pressures, deeply ingrained gender roles, and increased exposure to sexual harassment and violence. These factors can worsen feelings of depression and lower self-esteem. Adding to the complexity of the mental health landscape is the significant impact of age. Adolescents and young adults, particularly those aged 18-24, experience higher levels of stress, anxiety, and depression. This can be attributed in part to the pressures of academics, uncertainties about their future careers, and the difficulties of forming their identities (15). In addition, this specific group is more vulnerable to the negative impacts of social media, which can amplify feelings of inadequacy and diminish self-confidence.

Moreover, the effect of occupation on mental health is of great importance, as students and young professionals often experience high levels of stress and anxiety. This is influenced by the competitive nature of educational environments and the demands of the job market (16). The unrelenting drive for success and accomplishment creates an atmosphere that can lead to chronic fatigue, trouble focusing, and a decreased interest in activities. The pursuit of external validation, especially through the lens of social media, has been closely linked to various mental health challenges. The endemic comparison culture propagated by platforms such as Instagram and Facebook fosters negative life thoughts, low self-esteem, and feelings of inadequacy (17). Constantly seeking validation can lead to a harmful cycle of excessive eating or loss of appetite, disrupted sleep, and constant exhaustion, as people struggle to meet societal expectations. Delving deeper into the role of social media, the behaviors associated with posting and engagement on these platforms have been identified as significant contributors to mental health issues among Indian youths. The obligation to maintain a positive online persona has been linked with increased anxiety and depression, as it places undue stress on individuals and fosters feelings of insufficiency (18). Furthermore, excessive social media use can lead to disrupted sleep patterns, exacerbating problems related to sleep and contributing to a state of constant fatigue.

Given the backdrop, there is a need for targeted interventions that address the specific needs of different demographic groups and the importance of promoting healthy social media use patterns. Thus, the question underlining this research was:

- Which of these factors (gender, age group, occupation, external validation, and social media posting) influence mental health indicators such as negative life thoughts, low interest in activities, low self-esteem, poor appetite or overeating, trouble sleeping, feeling depressed, trouble concentrating, and constant fatigue among Indian youths?

To address the research question, we outlined, a clear and specific research objective statement as follows:

- To investigate and identify the factors that influences negative mental health indicators among Indian youths.

Methodology

Data Collection: Data for this research was systematically gathered using a quantitative approach. Among these participants, 48 are male and approximately 52% are female. The survey, structured with closed-ended questions, was conducted online to ensure broad and efficient reach among the targeted demographic. This method supports the collection of clear, quantifiable insights into the prevalence and determinants of negative mental health indicators (19).

Research Design: The study will adopt a cross-sectional survey design to assess the mental health indicators across different demographic and behavioral factors among Indian youths. The use of a cross-sectional design allows for the analysis of data at a single point in time, providing a snapshot of the mental health status of Indian youths in relation to gender, age group, occupation, external validation, and social media usage patterns.

Population and Sampling: The target population for this study will include 151 Indian youths aged 15-35, spanning various genders, age groups, occupational statuses, and levels of social media engagement. Although initially aiming for random sampling, low participation rates led to the adoption of convenience sampling to ensure sufficient data collection. This group comprises respondents from various parts of India, providing a broad perspective on the impact of social media across different demographic segments. Confidentiality was strictly maintained, with no personal identities collected or disclosed, ensuring the privacy of all participants.

Ethical Considerations: Ethical standards were rigorously maintained throughout the research process. Informed consent was obtained from all participants prior to data collection, ensuring that they were fully aware of the study's purpose and their role in it (20). Additionally, strict measures were implemented to maintain the confidentiality and anonymity of the participant responses, thereby safeguarding personal information and promoting an ethical research environment.

Questionnaire: This survey aims to evaluate the influence of demographic and behavioral factors on mental well-being in individuals. It utilizes a variety of research studies to investigate factors such as the amount of time spent on social media each day, the extent of personal information shared, the effect of receiving "likes" on self-esteem, the frequency of posting, the level of comfort with online versus face-to-face interactions, and various mental health indicators including sleep quality, mood, and self-perception. Research by (21) and (22) highlights the significance of analyzing the amount of time individuals spend on social media and the subsequent impact it has on their sleep patterns and mental state, respectively. Studies by (23) and (24) provide insights into the topics of personal information sharing and the psychological effects of receiving validation on social media. (25) and (26) offer valuable insights into the patterns of posting and the dynamics of social interactions on the internet compared to face-to-face interactions. The incorporation of mental health indicators is directed by established instruments such as the PHQ-9 and GAD-7 scales (27, 28), ensuring a thorough approach to comprehending the complex connection between social media usage and mental health results. The purpose of this questionnaire is to gather vital information about the intricate relationship between social media usage patterns and mental well-being, which will help in the creation of focused interventions.

To ensure data precision, our questionnaire incorporates various response scales to assess the demographic and psychographic variables ranges from "daily to no posting" (frequency of social media usage), "yes to not at all"(external validity), "employed-unemployed-student"(occupation), "15 to 30 and above"(age), "male to female"(gender).

Rigour and Trustworthiness: The rigour and trustworthiness of the study were upheld through meticulous design and implementation of the research methods. The reliability and validity of the instruments used for data collection were thoroughly evaluated to ensure that they accurately measure the intended variables (29). Consistency in data collection procedures and a robust data analysis plan contribute to the study’s credibility, allowing for reliable interpretation of the data collected (30).

Data Analysis Techniques: The methodology employed in this study encompasses both traditional statistical analyses and advanced machine learning techniques to examine the impact of social media on mental health among Indian youth. Initially, the study engaged in descriptive statistical analysis, focusing on categorical variables such as demographic and psychographic factors. This stage involved computing frequency distributions and percentiles to elucidate the sample's characteristics and identify patterns in the responses. Such a foundation was vital for understanding the data landscape, thereby setting the stage for more complex analyses.

For the primary analysis, non-parametric methods were utilized, specifically the Mann-Whitney U test and the Wilcoxon signed-rank test. These tests were selected due to their robustness in handling non-normally distributed data, a crucial aspect given the diversity of the survey responses. The study aimed to discern mean differences in mental health issues across various demographic and psychographic groups through these tests. A subsequent post hoc analysis employing the Wilcoxon method allowed for detailed pairwise comparisons between groups, shedding light on the intricate relationships between social media use and mental health outcomes across different segments of the population.

Model Selection: Enhancing the analytical rigor, the study incorporated two advanced machine learning algorithms. Firstly, the k-means clustering algorithm was applied (feature engineering), employing silhouette distance measures and the elbow method, to categorize the dataset into clusters based on eight identified mental health issues. This clustering facilitated a nuanced understanding of the dataset's inherent groupings.

Subsequently, the study utilized the CatBoost Classifier algorithm, renowned for its efficacy in handling categorical data classification. This algorithm aimed to identify significant features relative to the clusters as the target variable (identified using K means clustering), with the remaining five categorical variables serving as independent variables. The integration of the CatBoost algorithm enabled a thorough cross-validation of the results, bolstering the study's findings and offering a comprehensive view of the interplay between social media usage and mental health among Indian youth. Principal Component Analysis (PCA) was also employed to visually substantiate the clustering results.

Performance evaluation of our models revealed a CatBoost accuracy of 76.5%, demonstrating robust predictive power. The moderate silhouette score of 0.359, coupled with the Davies-Bouldin Index of 1.101 and Calinski-Harabasz Index of 110.182, confirmed the effectiveness of our clustering approach. These metrics collectively underscore the analytical rigor of our study, enhancing our insights into the relationships between social media usage and mental health.

Results

This study's demographic analysis reveals a nearly even gender distribution among the 151 respondents, with 52% female and 48% male. As shown in Table 1, the majority of respondents, 60%, post occasionally on social media, while 18% post twice a week, 13% daily, and 9% do not post at all. The age group distribution shows a skew towards younger individuals, with 43% in the 19-24 age bracket and 28% in the 25-30 range. Those above 30 years constitute 17%, and those below 18 are 13%. Regarding external validation from social media, 38% of respondents seek it a little bit and another 38% not at all, while 24% affirmatively seek validation. The data indicates a diverse age group and gender representation with varied frequencies of social media posting and different levels of reliance on external validation.

Table 2 reveals that participants who posted on social media twice a week exhibited the highest levels of mental health concerns, with an average mental score of 3.45, while those who didn't post at all showed the lowest concerns, scoring an average of 2.13. Individuals actively seeking external validation also reported higher negative mental health indicators, with an average score of 3.33, compared to those not seeking validation at 1.97. Occupation-wise, employed individuals reported more concerns than students or the unemployed. In terms of age groups, the below 18 category showed higher mental health concern scores, whereas those above 30 had lower scores. Gender-wise, females reported slightly higher levels of mental health concerns than males, with average scores of 2.70 and 2.44, respectively. This data highlights a correlation between social media habits, the need for external validation, occupation, age, gender, and various mental health concerns.

Table 3 reveals that the application of non-parametric tests offers a detailed understanding of the relationships between mental health concerns and various demographic and psychographic variables. Specifically, the Mann-Whitney U test, when applied to gender, resulted in a Chi-square value of 2.0904 with 1 degree of freedom and a p-value of 0.1482, suggesting no statistically significant difference in mental health concerns between genders, as the p-value exceeds the standard alpha level of 0.05. This indicates a similar distribution of mental health concerns among male and female respondents. The Kruskal-Wallis test, utilized for other variables, also demonstrates no significant variance in mental health concerns across different age groups (Chi-square = 4.2691, df = 3, p-value = 0.2338) and occupations (Chi-square = 2.1948, df = 3, p-value = 0.533), with p-values significantly higher than 0.05. Conversely, the need for external validation (Chi-square = 35.5671, df = 2, p < 0.0001) and frequency of social media posting (Chi-square = 30.0913, df = 3, p < 0.0001) demonstrate a significant association with mental health issues.

Table 1: Respondents Demographic and Psychographic Characteristics

Category	**Frequency**	**Percentage**
	Gender	
Male	73	48%
Female	78	52%
	Frequency of Post	
Daily	19	13%
No Post	14	9%
Occasionally	91	60%
Twice a week	27	18%
	Age Group	
19-24	65	43%
25-30	42	28%
Above 30	25	17%
Below 18	19	13%
	External Validation	
A little bit	57	38%
Not at all	58	38%
Yes	36	24%
	Occupation	
Employed	52	34%
Self-Employed	15	10%
Student	71	47%
Unemployed	13	9%
Total Respondents	151	100%

Table 2: Analysis of Mental Health Indicators Cross-tabulation across Demographic and Psychographic Variables among Respondents

Variable	**Category**	**Negative Life Thoughts**	**Low Interest in Activities**	**Low Self Esteem**	**Poor Appetite or Overeating**	**Trouble Sleeping**	**Feeling Depressed**	**Trouble Concentrating**	**Constant Fatigue**	**Average Mental Score**
Frequency of Post	Daily	2.63	2.58	2.79	3.05	3.00	3.32	3.26	3.42	3.01
	No Post	1.71	2.29	1.79	2.36	2.00	2.29	2.21	2.43	2.13
	Occasionally	2.21	2.15	2.18	2.20	2.38	2.32	2.27	2.58	2.29
	Twice a week	2.89	3.48	3.59	3.56	3.33	3.48	3.78	3.52	3.45
External Validity	A little bit	2.65	2.58	2.56	2.72	2.81	2.72	2.74	2.88	2.71
	Not at all	1.72	1.86	1.83	1.86	1.97	2.16	2.09	2.29	1.97
	Yes	2.83	3.22	3.36	3.44	3.28	3.33	3.47	3.67	3.33

Occupation	Employed	2.37	2.65	2.65	2.71	2.73	3.06	2.85	2.88	2.74
	Self-Employed	2.00	2.07	2.47	2.73	2.60	2.40	2.20	2.80	2.41
	Student	2.41	2.39	2.34	2.38	2.52	2.52	2.65	2.79	2.50
	Unemployed	2.23	2.46	2.46	2.77	2.46	2.00	2.54	3.00	2.49
Age Group	19-24	2.20	2.74	2.49	2.54	2.68	2.72	2.88	2.80	2.63
	25-30	2.52	2.31	2.33	2.71	2.83	2.62	2.62	2.74	2.59
	Above 30	1.92	2.00	2.28	2.32	2.08	2.40	1.88	2.72	2.20
	Below 18	2.95	2.42	2.95	2.63	2.47	2.79	3.05	3.37	2.83
Gender	Female	2.47	2.60	2.79	2.67	2.67	2.71	2.67	3.00	2.70
	Male	2.19	2.30	2.12	2.45	2.52	2.59	2.66	2.67	2.44

Table 3: Non-Parametric Test Results for Assessing Differences in Means across Categorical Variables if p value is less than 0.05 we can say there is a significant difference among or between the groups

Variables	Chi-square	DF	P value	Test
Gender	2.0904	1	0.1482	Mann-Whitney U test
Age Group	4.2691	3	0.2338	Kruskal-Wallis test
Occupation	2.1948	3	0.533	Kruskal-Wallis test
External Validation	35.5671	2	<.0001*	Kruskal-Wallis test
Frequency of Post	30.0913	3	<.0001*	Kruskal-Wallis test

Table 4: Post Hoc Analysis of Mental Health Indicators for Psychographic Variables

Variable	Level	Level	Score Mean Difference	Std Err Dif	Z	p-Value
Frequency of post	Twice a week	Occasionally	38.8771	7.481505	5.19643	<.0001*
	Twice a week	No Post	14.6971	3.934387	3.73555	0.0002*
	Occasionally	No Post	6.4698	8.714303	0.74243	0.4578
	Twice a week	Daily	4.1248	4.012046	1.02809	0.3039
	No Post	Daily	-6.0789	3.393849	-1.79117	0.0433*
	Occasionally	Daily	-18.2591	8.024021	-2.27556	0.0229*
External Validity	Yes	Not at all	31.5359	5.762988	5.47215	<.0001*
	Yes	A little bit	16.6784	5.738691	2.9063	0.0037*
	Not at all	A little bit	-24.454	6.198448	-3.94519	<.0001*

Table 4 provides a detailed post hoc examination of the significant differences in pairwise comparisons for variables that demonstrated substantial variation in previous results. This analysis employs the mean difference, standard error, Z-score, and p-value to assess the statistical significance between different levels of each variable. For the frequency of posts, the comparison between 'Twice a week' and 'Occasionally' posting reveals a mean difference of 38.8771 with a standard error of 7.481505, resulting in a Z-score of 5.19643 and a highly significant p-value of less than 0.0001. This indicates a substantial difference in mental health scores between these two posting frequencies. Similarly, 'Twice a week' versus 'No Post' shows a significant difference (mean difference = 14.6971, Std Err Dif = 3.934387, Z = 3.73555, p-value = 0.0002). However, the comparison between 'Occasionally' and 'No Post', as well as 'Twice a week' and 'Daily', did not show significant differences (p-values of 0.4578 and 0.3039, respectively). In contrast, the 'No Post' versus 'Daily' and 'Occasionally' versus 'Daily' comparisons show statistically significant differences with p-values of 0.0433 and 0.0229, respectively, suggesting notable variations in mental health scores for these posting frequencies. Regarding the need for external validation, the comparison between 'Yes' and 'Not at all' shows a significant mean difference of 31.5359 (Std Err Dif = 5.762988, Z = 5.47215, p-value < 0.0001). The 'Yes' versus 'A little bit' comparison also indicates a significant difference (mean difference = 16.6784, Std Err Dif = 5.738691, Z = 2.9063, p-value = 0.0037). Additionally, the 'Not at all' versus 'A little bit' comparison demonstrates a notable difference (mean difference = -24.454, Std Err Dif = 6.198448, Z = -3.94519, p-value < 0.0001).

Validation of Results

Cluster analysis: In the study's validation phase (Fig.1), cluster analysis (K-means), specifically employing the Elbow Method and silhouette analysis, was instrumental in determining the optimal cluster count within the dataset. The Elbow Method, which identifies a point of diminishing returns in within-cluster variance reduction, revealed a significant inflection at two clusters, suggesting a natural bifurcation of the data. This finding was further substantiated by silhouette analysis (Table 5), a measure of cluster cohesion and separation, which also indicated two as the optimal number. A detailed examination of these clusters revealed a stark contrast in mental health profiles: Cluster 1, consisting of 76 individuals, displayed significantly higher mean scores across all mental health issues compared to Cluster 2, which comprised 75 individuals.

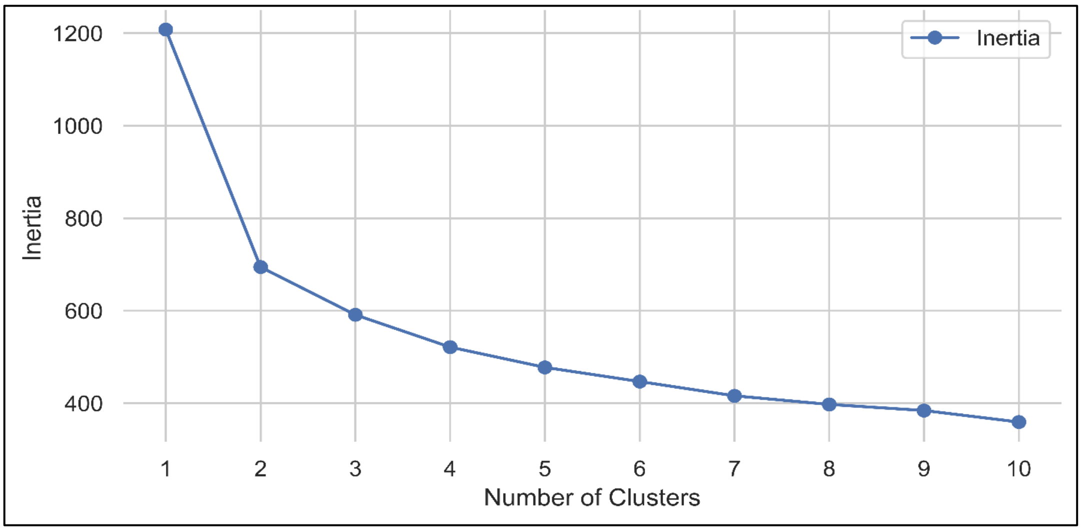

Figure 1: Elbow Method Diagram for Optimal Cluster Identification Using K Means Clustering

Table 5: Mental Health Issue Scores between Two Clusters

Variables	Cluster 1	Cluster 2
Trouble Sleeping	3.373333	1.828947
Low Interest in Activities	3.213333	1.710526
Feeling Depressed	3.573333	1.736842
Constant Fatigue	3.866667	1.828947
Poor Appetite or Overeating	3.52	1.618421
Low Self Esteem	3.466667	1.486842
Trouble Concentrating	3.6	1.736842
Negative Life Thoughts	3.293333	1.394737

The scatter plot represents the outcome of a K-means clustering algorithm applied to mental health issue data (Fig.2), with the spread of data points visualized through principal component analysis (PCA). Two distinct clusters are identified: Cluster 0 (red) and Cluster 1 (green), indicating the algorithm's grouping of individuals based on similarities in their mental health issue profiles. The overlap between the clusters suggests some commonality in mental health characteristics, whereas the areas of separation imply unique group-specific traits.

Feature importance: In applying a CatBoost classifier to evaluate the significance of various features in relation to social media-induced mental health issues (Fig.3), the analysis revealed a pronounced disparity in feature importance's. Notably, 'Frequency of Posts' emerged as the most influential factor, with an importance score of 57.27, suggesting a robust correlation between posting frequency and mental health outcomes. 'External Validation' was identified as the second most significant feature, with a score of 26.55, underlining the psychological impact of seeking social affirmation. In stark contrast, 'Occupation' registered no discernible influence, indicating its negligible role in the context of social media usage and mental health within this model.

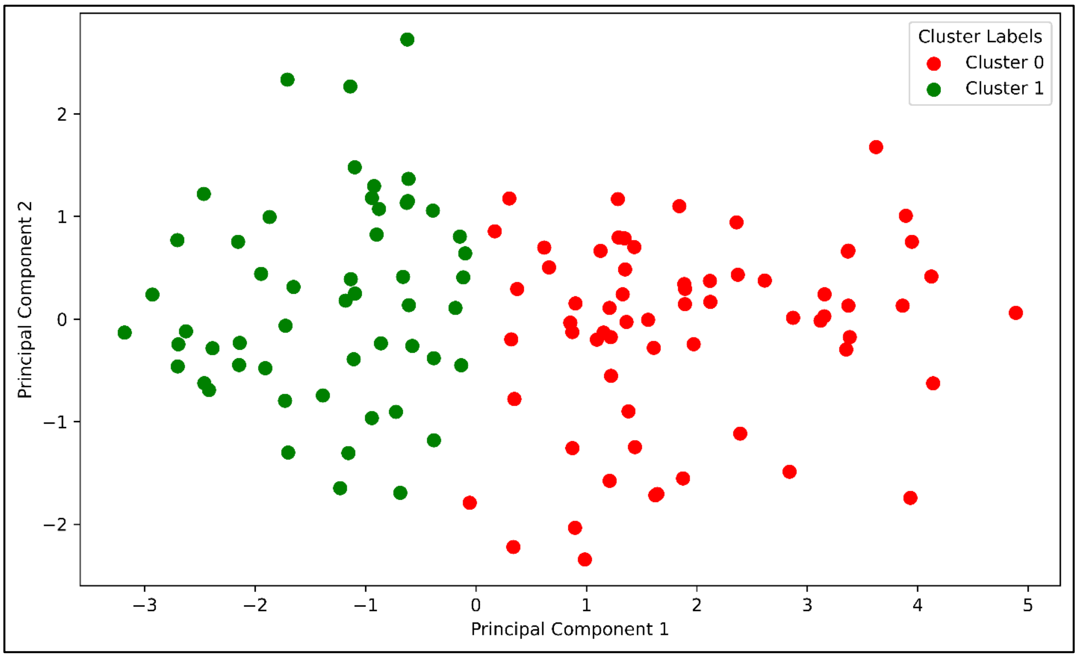

Figure 2: Principal Component Analysis of K-Means Clustering on Mental Health Issues

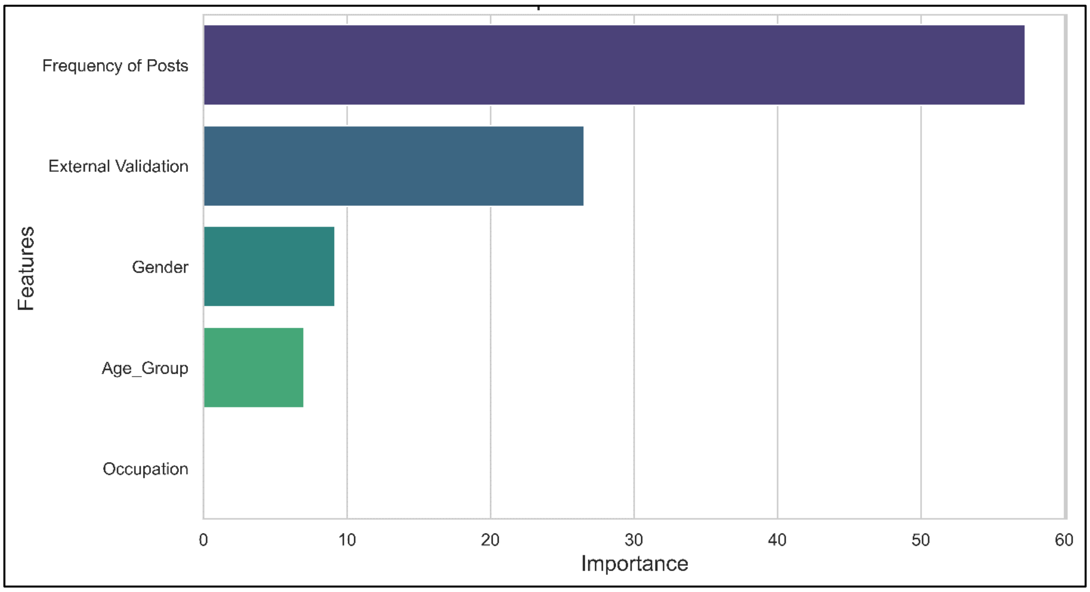

Figure 3: Feature Importance using CatBoost Classifier

Discussion

This research shows the significant influence of external validation and social media posting frequency on the mental well-being of Indian youths. It reveals a strong correlation with various mental health indicators, such as negative thoughts, decreased interest in activities, low self-esteem, poor appetite or overeating, trouble sleeping, feeling depressed, trouble concentrating, and constant fatigue. This demonstrates the importance of addressing the psychological impact of seeking validation and acknowledgment on social media platforms. The significance of these two factors, in comparison to gender, age group, and occupation, indicates that the digital environment has a crucial impact on the mental well-being of young individuals in India. A comprehensive strategy is needed, which involves teaching digital literacy, encouraging responsible social media use, and creating support networks to address the negative impacts of these influences. The findings highlight the importance of developing a resilient youth population capable of navigating external validation and social media pressures while safeguarding their mental health.

The literature review on mental health indicators among Indian youths highlights the complex interplay of various factors, including gender, age group, occupation, external validation, and social media posting. The study's findings underscore the significant influence of external validation and social media engagement in shaping mental health outcomes within these dimensions. Notably, gender disparities, as (14) elucidate, indicate that women may be more prone to depression and anxiety due to societal pressures and gender roles. However, their study goes beyond this by demonstrating how external validation and social media impact mental health across genders, revealing a universal risk. Similarly, age-related vulnerabilities, particularly in adolescents and young adults (15), are magnified in the digital era, where social media exacerbates stress from academic pressures and identity formation, highlighting the critical need for age-tailored interventions. Occupational stress, identified by (16) as a significant stressor due to competitive educational and job market pressures, is further compounded by digital interactions, suggesting a complex overlay of work-life and digital-life stressors. The literature's focus on the detrimental effects of social media on self-esteem and body image (6) finds parallel in the study's findings, which spotlight the significant influence of online validation seeking and social media posting behaviors on mental health outcomes, as supported by (18). This synthesis underscores the necessity for nuanced social media literacy and mental health education to mitigate the adverse effects of these digital behaviors. Consequently, although demographic factors such as gender, age, and occupation are important, this study highlights the significant impact of external validation and social media engagement. It suggests the need for interventions that address the wide range of ways social media affects the mental well-being of Indian youths, going beyond traditional demographic boundaries.

The study's findings support the conceptual framework based on Social Cognitive Theory, emphasizing the significant influence of external validation and frequency of social media posting on mental health indicators among Indian youths. These indicators include negative life thoughts, low interest in activities, and low self-esteem, among others. The connection between observational learning and self-efficacy is clearly seen in (11), where the pursuit of likes, comments, and shares for external validation significantly influences individuals' self-worth and mental health, corroborating the insights provided by (13) on the digital age's emphasis on external validation. Moreover, the frequency of social media posting, a behavior learnt through the lens of observational learning, plays a critical role in shaping mental health outcomes, with the nature of engagement and content consumed either exacerbating or alleviating mental health issues, echoing (12)'s observations on social media's impact. The complex relationship between personal factors (such as gender, age, and occupation) and environmental factors (including external validation and social media behavior) highlights the importance of taking a multi-dimensional approach in both research and interventions. In order to effectively address the impact of social media on the mental health of Indian youths, it is important to implement interventions that not only reduce negative effects but also encourage positive online experiences. This supports the framework's emphasis on the importance of targeted strategies to improve mental health outcomes in this demographic.

The study's findings emphasize the importance of external validation and social media posting frequency on mental health indicators among Indian youths. This calls for a reassessment of the mental health landscape within this demographic, especially in relation to digital interactions. The importance of external validation is highlighted as society increasingly depends on digital affirmation to boost self-esteem and shape identity. This aligns with the findings of (13), who emphasized the significant impact of social approval. In addition, the complex connection between social media activity and mental well-being suggests that excessive posting can worsen mental health issues (12). This highlights the need to include digital literacy and resilience-building in mental health interventions. This approach seeks to address the negative impacts of social media by encouraging thoughtful participation and decreasing reliance on seeking approval from others. Therefore, it is important to develop a comprehensive approach in research and policy-making. This approach should focus on creating interventions that target both the symptoms and root causes of digital-related mental health issues. By doing so, we can promote the well-being and resilience of young people in the face of constant digital influences.

The adoption of machine learning techniques alongside traditional statistical analyses in examining the impact of social media on mental health among Indian youth is highly relevant and advantageous. Machine learning offers sophisticated algorithms capable of detecting complex, non-linear patterns and interactions within large datasets, which are common in social media studies. This enables a deeper understanding of the nuanced effects of social media usage that traditional methods might miss (31). Additionally, machine learning can enhance predictive accuracy and provide insights into predictive factors of mental health issues, facilitating targeted interventions (32). By integrating both methodologies, this study can achieve a more comprehensive and precise analysis, ensuring robust and actionable findings that can better inform policy and individual care strategies.

Conclusion

In sum, our research on mental health indicators among Indian youths illustrates the complex relationship between different factors and mental well-being. External validation and social media usage are found to significantly impact a range of issues related to mental health. It is important to recognize the complex connection between digital literacy and mental health. To address this, interventions should be tailored to incorporate digital literacy into mental health programs. The focus should be on building resilience and encouraging healthier social media habits to minimize the psychological effects of seeking validation online. The study highlights the need for a comprehensive approach to mental health care, promoting tailored support systems that consider the diverse experiences of various groups. It underscores the significance of internal validation over external validation. This research provides valuable insights into the mental health of Indian youths in the digital age. It emphasizes the importance of collaboration among policymakers, practitioners, and stakeholders to create supportive environments that promote mental well-being and resilience. In order to inform the development of culturally sensitive mental health interventions, further research into these dynamics is essential. The study's implications for future mental health strategies and policy-making are significant, emphasizing the need for comprehensive approaches. With respect to methodology of future studies, high ethical standards in future digital health studies should be maintained, particularly those examining mental health indicators among youths, several key principles should be followed. Obtaining informed consent is crucial, ensuring participants are fully aware of the study's scope, risks, and benefits in easily understandable language (20). It is also essential to protect participant confidentiality and anonymity through secure data handling and anonymized reporting of results (33). Researchers are also advised to employ advanced cybersecurity measures to safeguard data and maintain transparency regarding data use. Studies should be culturally sensitive and tailored to the participant's cultural context. According to the author, ethical review and oversight by independent bodies should be regular to adapt to emerging issues and ensure compliance with ethical standards. As argued, the welfare of participants should always be a priority, with mechanisms in place for participants to withdraw freely and access mental health support if needed. These practices not only uphold ethical standards but also enhance the validity of research findings and trust with participants.

Abbreviation

Nil

Acknowledgement

Nil

Authors' Contributions

All authors have equally contributed.

Conflicts of Interest

The authors declare no conflict of interest.

Ethics Approval

Ethical clearance for this study was based on the provision of informed consent from all participants, ensuring their voluntary participation and understanding of the research's purpose and procedures.

Funding

This research received no external funding.

References

1. Liu S, Yang L, Zhang C, Xu Y, Cai L, Ma S, et al. Gender differences in mental health problems of healthcare workers during the coronavirus disease 2019 outbreak. Journal of psychiatric research. 2021;137:393–400.
2. Van Droogenbroeck F, Spruyt B, Keppens G. Gender differences in mental health problems among adolescents and the role of social support: results from the Belgian health interview surveys 2008 and 2013. BMC psychiatry. 2018;18(1):1–9.
3. Walton CC, Rice S, Gao CX, Butterworth M, Clements M, Purcell R. Gender differences in mental health symptoms and risk factors in Australian elite athletes. BMJ Open Sport—Exercise Medicine. 2021;7(1).
4. Sedgewick F, Leppanen J, Tchanturia K. Gender differences in mental health prevalence in autism. Advances in Autism. 2020;7(3):208–24.
5. Ferreira HG. Gender differences in mental health and beliefs about Covid-19 among elderly internet users. Paidéia (Ribeirão Preto). 2021;31.
6. Gao T, Mei S, Li M, D'Arcy C, Meng X. Decision authority on positive mental health in the workforce: A moderated mediation model of social support, gender, income, and occupation. Journal of Happiness Studies. 2021;1–17.
7. Thiruvenkatachari B, Sivakumar P, Ananth S, Sabbagh Y, Lewis BR, Chadwick SM, et al. The impact of COVID-19 pandemic on orthodontic services and trainees' mental health in India. Frontiers in Medicine. 2023;10.
8. Abas N, Hussin H, Mohd Hardi N, Hashim N. Exploring the interconnection of social media, mental health and youth: A bibliometric analysis. Social and Management Research Journal (SMRJ). 2023;20(2):185–206.
9. Jungari S. Maternal mental health in India during COVID-19. Public Health. 2020;185:97.
10. Wandrekar JR, Nigudkar AS. What do we know about LGBTQIA+ mental health in India? A review of research from 2009 to 2019. Journal of Psychosexual health. 2020;2(1):26–36.
11. Bandura A. Social foundations of thought and action. Englewood Cliffs, NJ. 1986;1986(23–28).
12. Lup K, Trub L, Rosenthal L. Instagram# instasad?: exploring associations among instagram use, depressive symptoms, negative social comparison, and strangers followed. Cyberpsychology, Behavior, and Social Networking. 2015;18(5):247–52.
13. Valkenburg PM, Peter J, Schouten AP. Friend networking sites and their relationship to adolescents' well-being and social self-esteem. CyberPsychology & behavior. 2006;9(5):584–90.
14. Özdin S, Bayrak Özdin Ş. Levels and predictors of anxiety, depression and health anxiety during COVID-19 pandemic in Turkish society: The importance of gender. International journal of social psychiatry. 2020;66(5):504–11.
15. Feng Y, Ma Y, Zhong Q. The relationship between adolescents' stress and internet addiction: A mediated-moderation model. Frontiers in psychology. 2019;10:2248.
16. Gerards R, Welters R. The impact of financial pressure on unemployed job search and labor market outcomes. 2017;
17. Gao J, Zheng P, Jia Y, Chen H, Mao Y, Chen S, et al. Mental health problems and social media exposure during COVID-19 outbreak. Plos one. 2020;15(4):e0231924.
18. Farhat K, Aslam W, Sanuri S. Exploring Drivers and Outcomes of Lurking Engagement and Posting Engagement in Social Media-Based Brand Communities:: Case of Hedonic vs. Utilitarian Brands. Asia Proceedings of Social Sciences. 2021;8(1):21–4.
19. Goddard W, Melville S. Research methodology: An introduction. Juta and Company Ltd; 2004.
20. Kapp MB. Ethical and legal issues in research involving human subjects: do you want a piece of me? Journal of clinical pathology. 2006;59(4):335–9.
21. Twenge JM, Campbell WK. Associations between screen time and lower psychological well-being among children and adolescents: Evidence from a population-based study. Preventive medicine reports. 2018;12:271–83.
22. Woods HC, Scott H. # Sleepyteens: Social media use in adolescence is associated with poor sleep quality, anxiety, depression and low self-esteem. Journal of adolescence. 2016;51:41–9.
23. Bayer JB, Ellison NB, Schoenebeck SY, Falk EB. Sharing the small moments: ephemeral social interaction on Snapchat. Information, communication & society. 2016;19(7):956–77.

24. Vogel EA, Rose JP, Roberts LR, Eckles K. Social comparison, social media, and self-esteem. Psychology of popular media culture. 2014;3(4):206.
25. Sheldon P, Bryant K. Instagram: Motives for its use and relationship to narcissism and contextual age. Computers in human Behavior. 2016;58:89–97.
26. Moreau A, Laconi S, Delfour M, Chabrol H. Psychopathological profiles of adolescent and young adult problematic Facebook users. Computers in Human Behavior. 2015;44:64–9.
27. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. Journal of general internal medicine. 2001;16(9):606–13.
28. Spitzer RL, Kroenke K, Williams JB, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Archives of internal medicine. 2006;166(10):1092–7.
29. Morse JM, Barrett M, Mayan M, Olson K, Spiers J. Verification strategies for establishing reliability and validity in qualitative research. International journal of qualitative methods. 2002;1(2):13–22.
30. Taylor SJ, Bogdan R, DeVault ML. Introduction to qualitative research methods: A guidebook and resource. John Wiley & Sons; 2015.
31. Elhai JD, Montag C. The compatibility of theoretical frameworks with machine learning analyses in psychological research. Current Opinion in Psychology. 2020;36:83–8.
32. Hornstein S, Forman-Hoffman V, Nazander A, Ranta K, Hilbert K. Predicting therapy outcome in a digital mental health intervention for depression and anxiety: A machine learning approach. Digital Health. 2021;7:20552076211060659.
33. Arifin SRM. Ethical considerations in qualitative study. International journal of care scholars. 2018;1(2):30–3.