A Report on

"THE RELATIONSHIP BETWEEN SOCIAL MEDIA ADDICTION, PSYCHOLOGICAL FACTORS AND SOCIAL FACTORS"

Submitted to

DEPARTMENT OF STATISTICS

SHIVAJI UNIVERSITY

KOLHAPUR

FOR THE PARTIAL FULFILMENT OF DEGREE OF MASTER OF SCIENCE IN STATISTICS

Submitted by

MR. PATIL SHUBHAM MADHUKAR

M.Sc. IInd (Statistics)

Under the Guidance of

PROF. DR. S. B. MAHADIK

NOVEMBER 2024

SHIVAJI UNIVERSITY, KOLHAPUR

DEPARTMENT OF STATISTICS

CERTIFICATE

This is to certify that **Mr. Patil Shubham Madhukar**, with **PRN: 2023000327**, a student of **M.Sc. (Statistics), Part II**, has successfully completed the research project work for **Semester III** of the academic year **2024-2025**.

The research was undertaken under the esteemed guidance of **Dr. S. B. Mahadik** as a mandatory component for the partial fulfilment of the degree requirements of **Master of Science in Statistics.**

Date:

Place:

Project Guide
Prof. (Dr.) S. B. Mahadik.

Prof. (Dr.) S. B. Mahadik.
Head of the Department,
Department of Statistics.

ACKNOWLEDGMENT

It gives me great pleasure to express my heartfelt gratitude to the Head, **Prof. (Dr.) S. B. Mahadik**, Department of Statistics, Shivaji University, Kolhapur, for their invaluable guidance, unfailing patience, and constant encouragement throughout the course of my project. Their insightful suggestions and expertise were pivotal in shaping the direction of this research.

I am also deeply indebted to all the faculty members of the Department of Statistics for their constructive feedback and support, which enriched my learning experience. A special mention to the non-teaching staff for their assistance in providing access to essential resources and facilities. I would like to extend my gratitude to my peers and friends, whose moral support and insightful discussions helped me overcome challenges during the project.

This project has been an incredible journey, and the knowledge and skills I have gained, both analytical and personal, will undoubtedly benefit me in my future academic and professional endeavors.

Mr. Shubham Madhukar Patil

Contents

1. Introduction

1.1 Background

Social media has become a central part of daily life, changing the way of peoples communicate, share information, and connect with others. While these platforms offer benefits, such as easier communication and access to information, they also come with concerns, particularly regarding addiction. Social media addiction refers to the compulsive and excessive use of platforms like Facebook, Instagram, WhatsApp, Snapchat etc. Social media addiction can negatively affect both mental health and social connections. Psychologically, it can lead to increased anxiety, depression, and lower self-esteem, as people often compare themselves to others online or seek constant validation through likes and comments. This can create a cycle of insecurity and emotional distress. Socially, addiction to social media may harm real-life relationships, as people spend more time online and less time engaging with friends and family in person. It can also contribute to feelings of isolation, as virtual interactions replace deeper, more meaningful connections. Overall, excessive use of social media can disturb both mental well-being and social life.

1.2 Research problem

In recent years, the widespread use of social media has raised concerns about its impact on mental health and social well-being. Many individuals exhibit behaviors associated with social media addiction, which can disrupt daily functioning and relationships. While numerous studies have explored factors related to social media use, no one have examined how both psychological and social factors jointly influence the likelihood of developing social media addiction. This gap in understanding creates a need to investigate how specific social media addiction relate with psychological factors and social factors. Addressing this problem can provide insights for mental health interventions and promote healthier social media use habits. To check the relations, we select seven psychological factors such as Self-esteem, Depression, Anxiety, *Fear of Missing Out* (FOMO), Social comparison, Emotional regulation and Attachment style and also, we select five social factors such as Family Structure, Social network, Peer groups, cultural background and Education. With these selected factors we checked the how social media addiction shows relationship with each factor.

1.3 Objective

“To explore the relationship between social media addiction, psychological factors and social factors among **Under Graduate Students** in Kolhapur City.”

1.4 Scope

As social media has become an integral part of young adults' daily routines, it significantly impacts their mental health, academic performance, and social interactions. This research seeks to identify specific psychological factors, including stress, anxiety, self-esteem, and emotional well-being, that may increase susceptibility to social media addiction. Additionally, it will investigate social factors, such as peer influence, family relationships, social support, and sense of community, to understand how these aspects shape addictive behaviors. By focusing on Under Graduate students, the study addresses a critical developmental stage, as they are forming social connections and identities, which may heighten vulnerability to excessive social media use. The scope is intentionally limited to this group to provide precise insights into a high-risk population that is frequently targeted by social media marketing. Findings from this research will assist educators, mental health professionals, and policymakers in designing effective interventions to mitigate social media dependency among students. Ultimately, this study aims to promote a balanced and healthier approach to social media use within academic and social communities

2. Key Terms

2.1 Social Media Addiction

Social media addiction occurs when a person becomes excessively dependent on social media platforms, often spending more time online than they intend to. This compulsive use can interfere with daily activities, relationships, and responsibilities, leading to a decline in productivity and well-being. People struggling with social media addiction may experience difficulty controlling the amount of time spent on these platforms, and they may feel anxious, frustrated, or upset when unable to access social media. The addiction is typically measured using scales like the ***Bergen Social Media Addiction Scale*** **(BSMAS)**, which assesses various dimensions of social media use, including salience, tolerance, and withdrawal symptoms.

2.2 Psychological Factors

i. **Self-esteem** plays a crucial role in social media addiction, especially for individuals who may have lower self-worth. Researchers suggests that people with low self-esteem often seek external validation to feel good about themselves. Social media platforms provide immediate feedback through likes, comments, and shares, which can become a way to gain temporary boosts in self-esteem.

ii. **Depression** is a key psychological factor linked to social media addiction, as individuals often use social media to escape negative emotions or combat isolation. While it provides temporary relief, constant exposure to idealized content can worsen feelings of loneliness and inadequacy. This creates a cycle where increased social media use exacerbates depression, leading to addiction. Studying this connection among under graduate students is crucial due to their vulnerability during this developmental stage.

iii. **Anxiety:** Social media offers temporary relief for anxiety by providing connection and reassurance. However, excessive use can intensify anxiety due to information overload, negative comparisons, and the pressure to maintain an idealized online image. Compulsive engagement, driven by the need for validation through likes and comments, often creates a cycle where anxiety fuels social media use, which in turn worsens anxiety. This cycle is especially harmful to vulnerable groups like students, who already face academic and social pressures.

iv. **FOMO,** or "Fear of Missing Out," is a psychological phenomenon where individuals feel anxious or excluded when not part of trends or activities showcased on social media. Platforms amplify this by displaying others' experiences, driving compulsive checking and increased screen time. This cycle fosters dependency as individuals seek validation through likes and comments or feel pressured to stay constantly connected. FOMO is particularly prevalent among students, who are more prone to comparisons and anxiety about missing out on social or cultural events, contributing significantly to social media addiction.

v. **Social Comparison,** the act of evaluating oneself against others, is intensified on social media, where users encounter curated highlights of others' achievements, relationships, and appearances. This often triggers feelings of inadequacy, anxiety, or jealousy, leading individuals to seek reassurance or approval online. The cycle of comparison and validation reinforces social media addiction, as users spend more time online to gain positive feedback or escape negative emotions. Over time, this reliance on social comparison can harm mental health and deepen addictive behaviors. Addressing this dynamic is critical, especially for vulnerable groups like students, who are more susceptible to its effects.

vi. **Emotional regulation,** the ability to manage emotions effectively, is closely linked to social media addiction. Many individuals use social media to escape feelings of stress, loneliness, or sadness, seeking quick rewards like likes and comments for temporary relief. However, this reliance can lead to avoiding real problems and create a cycle of dependency, where social media becomes a go-to solution for negative emotions. Those with difficulty managing their emotions are particularly prone to this pattern, increasing their risk of addiction. This connection is especially significant for students, who may be more vulnerable to using social media as an emotional escape.

vii. **Attachment styles**, shaped by early relationships, can influence social media addiction. Individuals with anxious attachment may overuse social media for reassurance and validation, while those with avoidant attachment may prefer it for low-pressure connections. Social media provides a quick sense of connection but often fosters dependency, as people with insecure attachment styles rely on it to meet unmet emotional needs, increasing the risk of addiction

2.3 Social Factors

i. **Family structure** impacts social media addiction. Individuals from single-parent or high-conflict families may turn to social media for connection, while those from supportive families are less dependent on it. Parental involvement in setting limits can prevent excessive use, particularly in younger individuals. A positive family environment fosters healthy boundaries, while less supportive families may increase social media reliance, influencing whether it supplements or replaces real-life connections.

ii. **Social network,** while contributing to social media addiction, can also have positive effects. They help individuals stay connected with distant friends and family, fostering a sense of community and belonging. Social media enables new connections based on shared interests, offering opportunities for personal, professional, and academic growth. It provides emotional support, reduces isolation, and can be a source of motivation and positive content. When used in moderation, social networks enhance communication and relationships.

iii. **Peer groups** significantly influence social media addiction, particularly among younger individuals. Students, seeking validation from peers, may feel pressured to stay active on social media to fit in. As peers share posts, update statuses, and engage in online trends, others may feel compelled to do the same, increasing their social media usage. This can create competition and comparison, leading to feelings of inadequacy and dependency on social media for approval. Peer groups often use social media to stay connected, making it harder to disconnect without feeling left out. Over time, this cycle can contribute to addiction, prioritizing online interactions over real-life connections.

iv. **Cultural background** plays a key role in social media use and addiction. In cultures where online presence is tied to social status and peer approval, individuals may feel pressured to spend more time on social media to maintain a perfect online image. In cultures valuing community and group identity, social media becomes a tool for connection and participation in cultural activities. Conversely, in individualistic cultures, social media may be used for self-expression and personal success, leading to overuse. Cultural factors such as family values, media habits, and social interaction norms influence social media usage and the potential for addiction, making it essential to consider these influences when addressing the issue.

v. **Education** influences social media use and the risk of addiction. Students, particularly in higher education, use social media to stay connected with peers, receive academic updates, and engage in school activities. However, excessive use can distract from studying and academic focus. Schools also contribute to social media use by incorporating online platforms for learning and communication. Students facing academic stress may turn to social media as a coping mechanism, increasing dependence. Educating students about the risks of overuse and teaching time management skills are crucial to preventing social media addiction.

3. Literature Review

The relationship between social media use and its psychological impact has garnered increasing attention in recent years, particularly as social media becomes more integrated into daily life. Numerous studies have examined the addictive nature of social media and its effects on various aspects of mental well-being, including happiness, life satisfaction, and overall psychological health.

Necmettin Cifcti, Metin Yildiz et.al (2023)

This study investigates the relationship between social media addiction, happiness, and life satisfaction among 1008 adults in Turkey. It finds that increased social media addiction correlates with decreased levels of happiness ($\beta=-0.290$, $p<0.05$) and life satisfaction ($\beta=-0.235$, $p<0.05$). Life satisfaction mediates the effect of social media addiction on happiness ($\beta=-0.139$). The best predictive model for happiness was identified as elastic net regression, highlighting the importance of life satisfaction and social media addiction in predicting happiness levels.

Daniel zarate, Evita March et.al (2023)

The study investigates the psychometric properties of the Bergen Social Media Addiction Scale (BSMAS) using Item Response Theory (IRT) in a sample of 968 participants. It finds that higher scores on the BSMAS indicate a greater risk of problematic social media use (PSMU), particularly among younger individuals and females. The analysis reveals significant item discrimination and reliability, with a proposed cut-off score of 26 for identifying high-risk individuals. Overall, the BSMAS is deemed a reliable measure for assessing PSMU.

Shivani Arora, Manoj Kumar et.al (2022)

The study investigates social media addiction in India using the Bergen Social Media Addiction Scale (BSMAS), revealing that 18% of respondents are at risk, with males being slightly more affected than females. The research highlights those younger individuals, particularly those aged 21-30, are more prone to addiction, and undergraduates show higher risk levels compared to undergraduates. The findings emphasize the need for awareness and preventive measures in educational institutions to address this growing concern.

4. Methodology

4.1 Study Design

This study uses a quantitative and qualitative research design to explore how social media addiction is linked to psychological and social factors in under graduate students. The main goal is to understand how factors like self-esteem, depression, anxiety, FOMO, social comparison, emotional regulation, and attachment style influence social media addiction. It will also examine how social factors such as family structure, social networks, peer groups, cultural background, and education impact social media use. By focusing on under graduate students, the study aims to look at how this age group, which is highly active on social media, is influenced by these factors.

The study will use a survey-based method for data collection. The survey will consist of a custom-made questionnaire that includes 27 questions related to psychological factors, which are divided into 7 factors: self-esteem, depression, anxiety, FOMO, social comparison, emotional regulation, and attachment style. These factors are important because past research has shown that people with lower self-esteem, more anxiety, and a tendency to compare themselves to others often become more dependent on social media.

In addition to the psychological factors, the questionnaire will also include 19 questions on social factors, divided into 5 factors: family structure, social network, peer groups, cultural background, and education. These social factors matter because people's social environments influence how they use social media. For example, family dynamics and peer pressure can drive social media use, and education can affect how aware someone is about the potential effects of social media.

To measure social media addiction, the study will use the "Bergen Social Media Addiction Scale (BSMAS)", which includes 6 questions. Each question uses a 5-point

Likert scale, asking how often a person experiences behaviors like checking social media excessively or feeling anxious when unable to check it. The BSMAS is a well-established and reliable tool used in previous studies to measure social media addiction.

The survey will be distributed to undergraduate students using convenience sampling, which means selecting participants who are easily accessible. The goal is to get a diverse group of students from various universities. Once the data is collected, it will be analyzed using statistical methods.

For the data analysis, the responses to the Likert scale questions will be converted into numerical values for statistical analysis. The scale used in the questionnaire ranges from 1 to 5, with 1 representing the lowest level of agreement or frequency and 5 representing the highest level. However, for the questions that are phrased negatively, the scores will be reversed. In these cases, a response of "1" will be converted to "5," and "2" will be converted to "4," and so on. This adjustment is made to ensure consistency in how the responses are interpreted across positively and negatively worded items, making sure that higher scores always reflect more intense levels of the construct being measured (whether that’s addiction, anxiety, or another factor)

4.2 Variables

Dependent Variable: Social Media Addiction (measured by BSMAS scores).
Independent Variables:

Psychological factors: Self-esteem, Depression, Anxiety, FOMO, Social Comparison, Emotional Regulation, Attachment style.

Social factors: Family Structure, Social network, Peer Groups, Cultural Background, Education.

4.3 Sampling and Data collection

For this research, a comprehensive list of all colleges in Kolhapur city was compiled. Each college name was written on individual cards, and a lottery method was used to randomly select colleges. The sample size was then determined through convenience sampling. To ensure the sample was representative, we maintained proportionality based on gender and academic streams (Arts, Commerce, Science, Engineering, Medical, Agriculture, Law). A total of 179 students, including 101 females and 78 males, were selected through this method across 12 colleges. Data collection was carried out via structured, face-to-face interviews, which facilitated the accurate capture of information regarding psychological and social behaviors, as well as social media usage patterns. Below table provides detailed information on the data collection process.

Sr. No	Stream	Male_Count	Female_Count	Total_Count
1	Arts	2 (9.09)	20 (90.91)	22
2	Commerce	6 (50)	6 (50)	12
3	Science	15 (55.56)	12 (44.44)	27
4	Engineering	20 (48.78)	21 (51.22)	41
5	Agriculture	12 (52.17)	11 (47.83)	23
6	Law	11 (47.83)	12 (52.17)	23
7	Medical	12 (38.71)	19 (61.29)	31
8	Total	78 (43.58)	101 (56.42)	179

Table 1: Stream wise count and percentage of Male & Female in Data

4.4 Data Analysis Techniques

4.4.1 Univariate Analysis

Descriptive Statistics: Descriptive statistics summarize and provide an overview of the key characteristics of the dataset, such as the mean, median, mode, standard deviation, and range. These statistics help to understand the central tendency and variability of individual variables.

Histogram and Boxplot: A histogram is used to visualize the distribution of a single variable, showing the frequency of data within certain ranges. A boxplot provides a graphical representation of the data's spread, identifying the median, quartiles, and potential outliers, allowing for an easy comparison of the distribution of the data.

4.4.2 Bivariate Analysis

Scatterplot: A scatterplot displays the relationship between two continuous variables, allowing for the visualization of trends, patterns, or correlations between them. It helps in identifying linear or non-linear associations and outliers in the data.

Correlation Test: This test is used to measure the strength and direction of the linear relationship between two continuous variables. The Pearson correlation coefficient (r) is commonly used, with values ranging from -1 to 1, indicating the degree of correlation.

Chi-square Contingency Table: A chi-square test is used to assess whether there is a significant association between two categorical variables. The contingency table displays the frequency distribution of the variables, and the chi-square test determines if the observed distribution differs from the expected.

Multiple Linear Regression: Multiple linear regression is used to model the relationship between a dependent variable and multiple independent variables. It helps to predict the value of the dependent variable based on the values of the independent variables and estimate their respective effects.

4.4.3 Statistical Model

Binomial Logistic Regression: Binomial logistic regression is used when the dependent variable is binary (i.e., two possible outcomes). It estimates the probability of an event occurring, based on one or more predictor variables, and produces odds ratios to interpret the influence of predictors on the outcome

4.4.4 Machine Learning Algorithms

Random Forest: Random Forest is an ensemble learning algorithm used for classification and regression tasks. It builds multiple decision trees during training and outputs the most frequent class (for classification) or the average prediction (for regression) from all trees, improving accuracy and reducing overfitting.

XGBoost: XGBoost (Extreme Gradient Boosting) is a machine learning algorithm that is highly efficient for classification and regression problems. It uses gradient boosting to optimize predictive accuracy, handling missing data and preventing overfitting through regularization techniques

These methods collectively allow for both exploratory and predictive analysis, providing comprehensive insights into the data from various angles.

5. Analysis and Results

5.1 Univariate Analysis

To understanding the nature of data we seen descriptive statistics, histogram and boxplot of individual variable. Here we use two types of datasets first is original dataset and another is normalized dataset.

5.1.1 Descriptive statistics

	Factor Name	Mean	Standard Deviation	Mean (Normalization)	Standard Deviation (Normalization)
Psychological Factor	Self Esteem	7.74	0.19	7.74	0.19
	Depression	10.99	0.26	8.80	0.21
	Anxiety	8.48	0.24	8.48	0.24
	FOMO	10.89	0.16	10.89	0.16
	Social Comparison	8.29	0.17	11.05	0.23
	Emotional Regulation	7.63	0.17	10.17	0.22
	Attachment Style	9.70	0.22	9.70	0.22
Social Factor	Family Structure	20.65	0.23	16.52	0.18
	Social Network	10.24	0.17	13.65	0.23
	Peer Groups	9.87	0.15	13.15	0.20
	Cultural Background	15.73	0.24	15.73	0.24
	Education	15.83	0.21	15.83	0.21
Social Media Addiction	BSMAS ALL	14.10	0.35	9.40	0.23

Table 2: Descriptive statistics of psychological, social factors and BSMAS.

Psychological Factors

1. **Highest Mean**: **Depression** (Mean: 10.99) and **FOMO** (Mean: 10.89) are the most prominent psychological factors, indicating these aspects may exert a stronger influence on individuals.
2. **Moderate Influence**: **Attachment Style** (Mean: 9.70) and **Anxiety** (Mean: 8.48) show moderate importance among psychological factors.
3. **Lowest Mean**: **Self-Esteem** (Mean: 7.74) and **Emotional Regulation** (Mean: 7.63) appear to have less impact compared to other psychological factors.

Social Factors

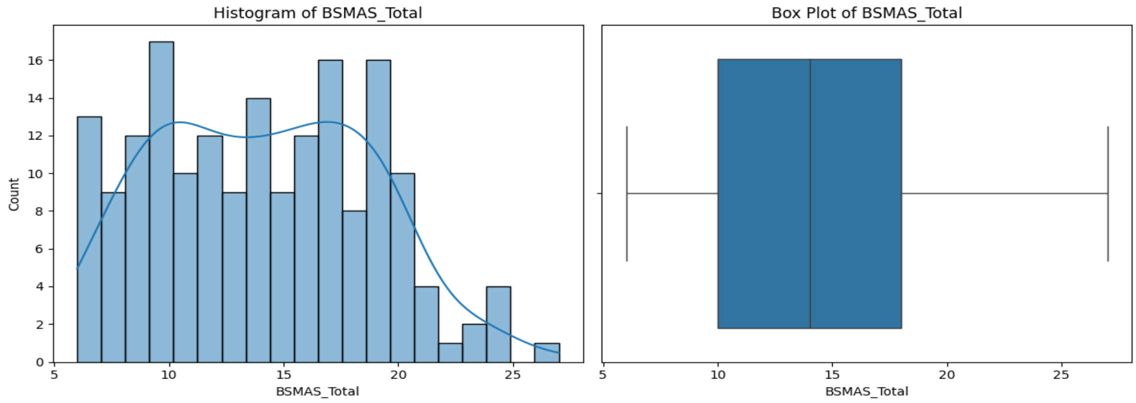
1. **Dominant Factors**: **Cultural Background** (Mean: 15.73) and **Education** (Mean: 15.83) emerge as the most significant social factors.
2. **Peer Influence**: **Social Network** (Mean: 10.24) and **Peer Groups** (Mean: 9.87) gain prominence, highlighting their growing importance compared to **Family Structure** (Mean: 20.65).
3. **Family Structure**: Despite a high raw score, its normalized mean suggests it is less critical than previously inferred.

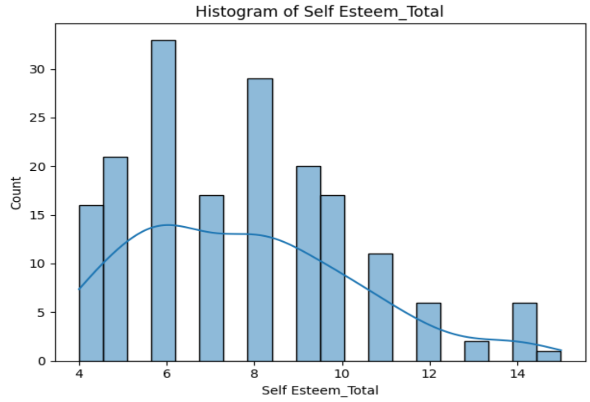
Social Media Addiction (BSMAS)

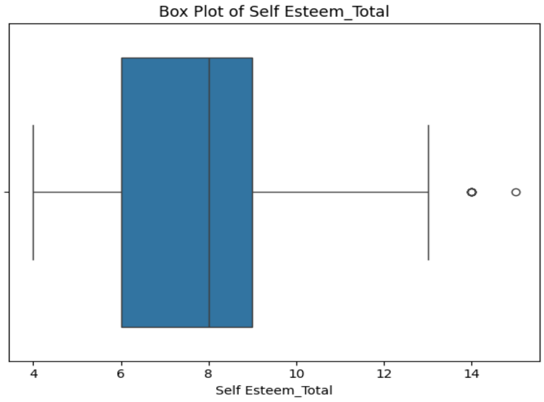
- The normalized mean of **9.40** suggests that social media addiction has a lower relative importance compared to psychological and social factors, despite its high variability.

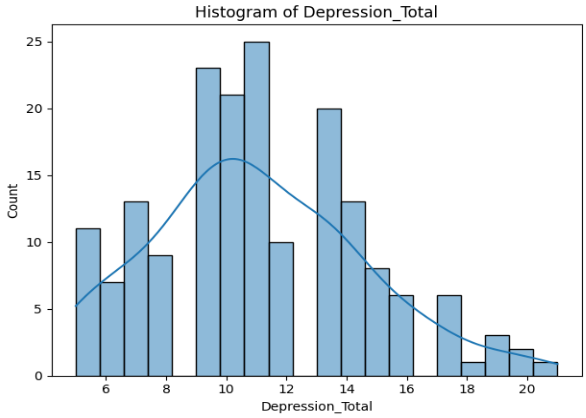
Normalization adjusts the means and standard deviations to a uniform scale, which makes comparisons across factors more meaningful. Changes in means (e.g., Depression, Social Comparison) suggest adjustments to align scales, allowing for better integration into a unified analysis.

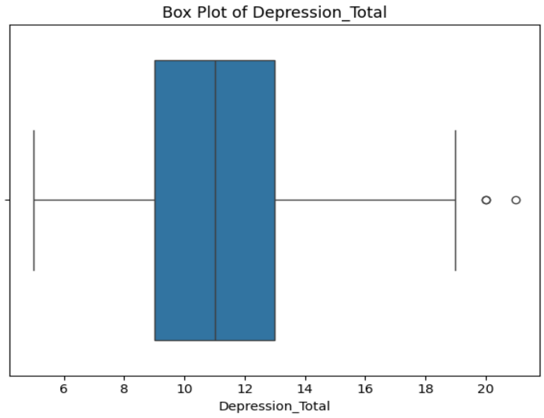
5.1.2 Histogram and boxplots for BSMAS, Psychological and Social Factors

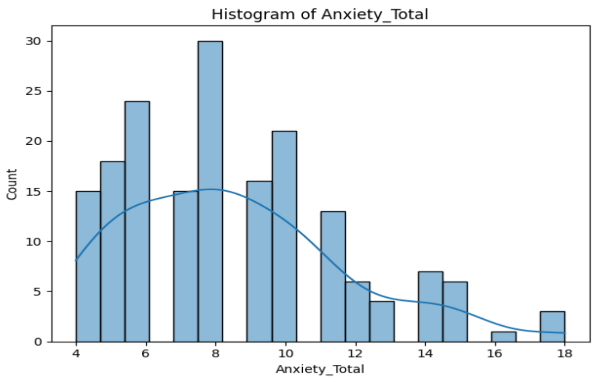


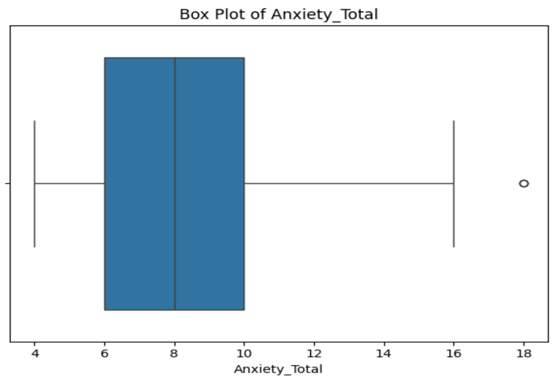

Histogram of Self Esteem_Total
Count
Self Esteem_Total

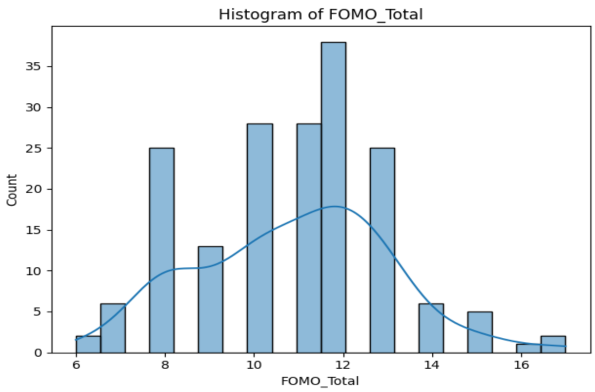

Box Plot of Self Esteem_Total
Self Esteem_Total

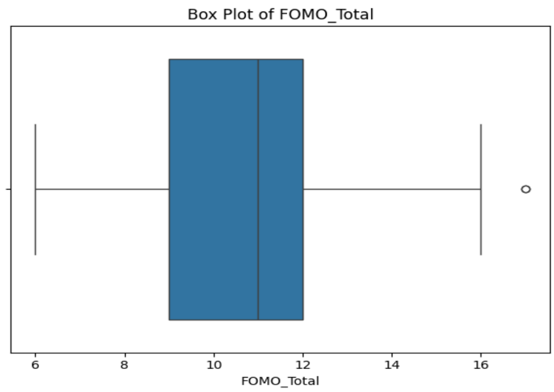

Histogram of Depression_Total
Count
Depression_Total

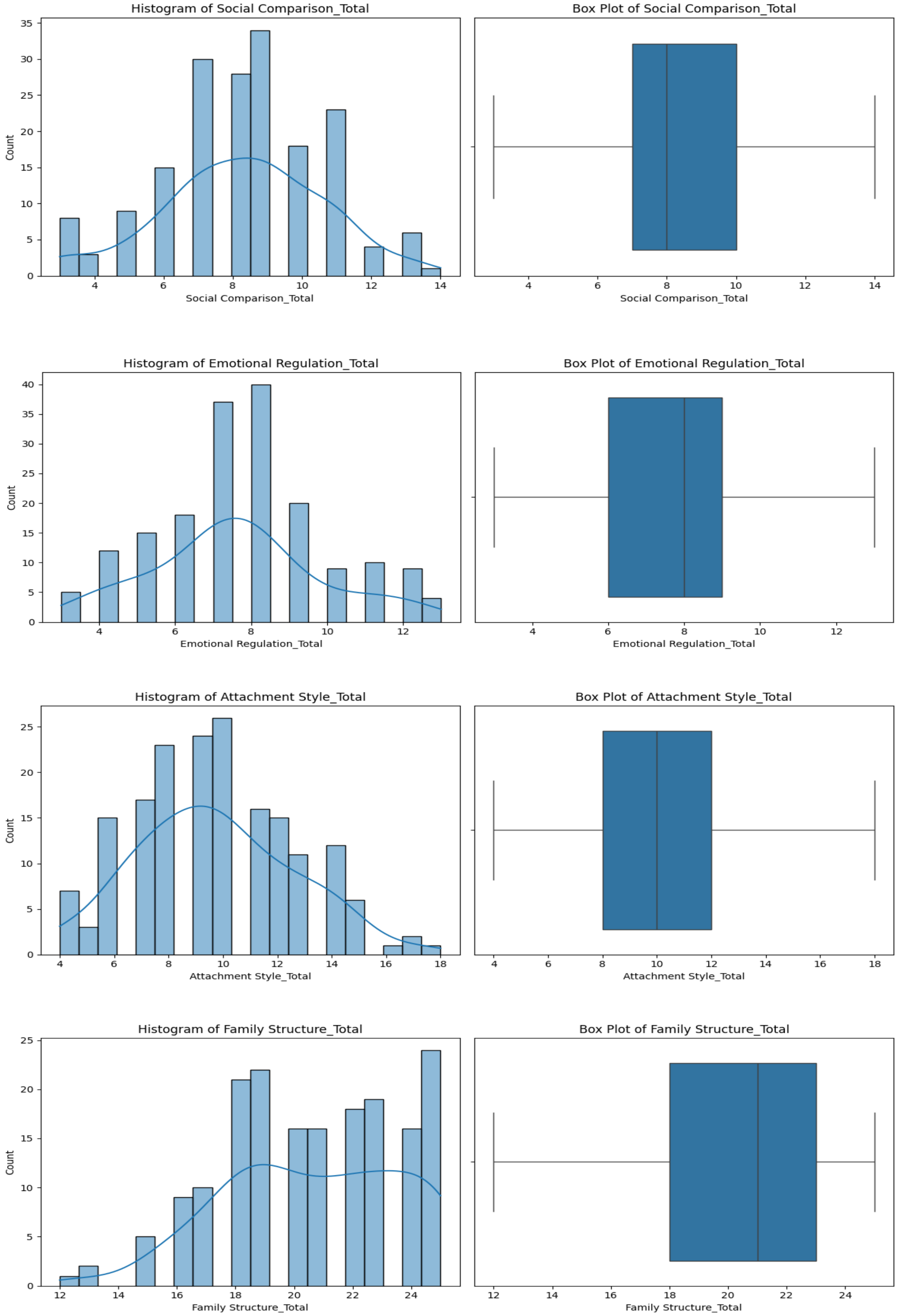

Box Plot of Depression_Total
Depression_Total


Histogram of Anxiety_Total
Count
Anxiety_Total


Box Plot of Anxiety_Total
Anxiety_Total


Histogram of FOMO_Total
Count
FOMO_Total


Box Plot of FOMO_Total
FOMO_Total


Histogram of Social Comparison_Total
Box Plot of Social Comparison_Total
Count
Social Comparison_Total
Histogram of Emotional Regulation_Total
Box Plot of Emotional Regulation_Total
Emotional Regulation_Total
Histogram of Attachment Style_Total
Box Plot of Attachment Style_Total
Attachment Style_Total
Histogram of Family Structure_Total
Box Plot of Family Structure_Total
Family Structure_Total

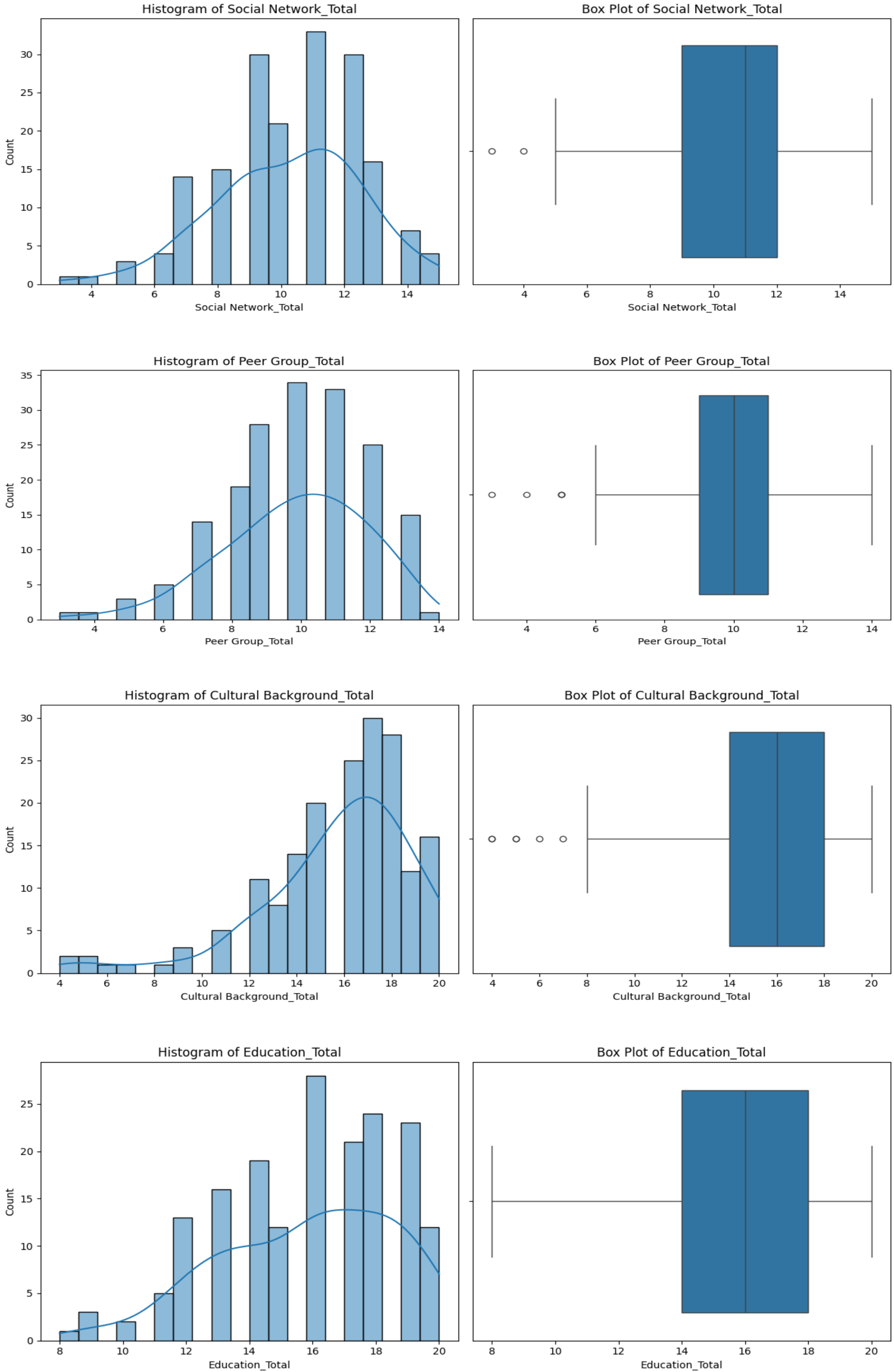


Fig 1. Histogram and boxplots for Likert Scale Data of BSMAS, Psychological and Social Factors

The BSMAS_Total scores suggest that most individuals exhibit moderate levels of social media addiction, as reflected by the median score around 14 and the concentration of scores between 10 and 18. While a small number of individuals exhibit higher levels of addiction (scores above 20), these cases are not common. The lack of outliers indicates the dataset is consistent, and the moderate variability reflects a reasonably diverse range of addiction levels.

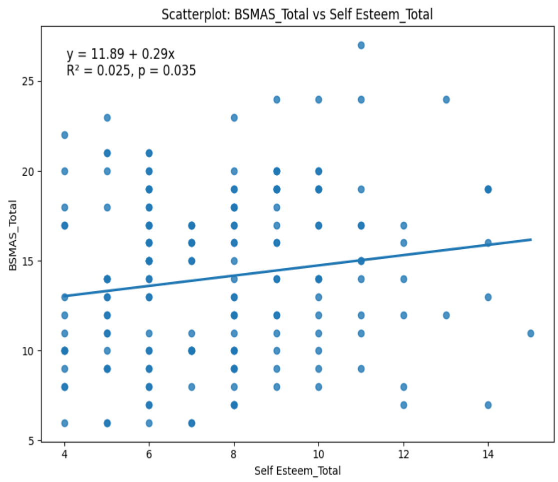
Generally, all psychological factors are left skewed in data. And social factors are right skewed. BSMAS score is also left skewed and distributed ranging from 5 to 20 score.

Boxplots of the variables shows the data distribution especially for median of the data. The outliers showed by boxplots are not actually outliers these are higher score observations for that respective factor.

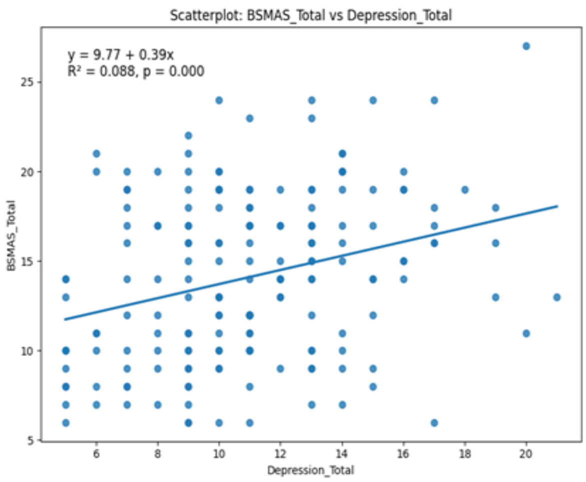
5.2 Bivariate Analysis

To understand the relationship between social media addiction and seven psychological factors and 5 social factors. We have plotted scatter plots with social media addiction and other factors. Pearson, Spearman and Kendall methods are used to check correlation. The multiple linear regression technique used to derive relationship between social media addiction and other factors. Each single factor is regressed with social media addiction score.

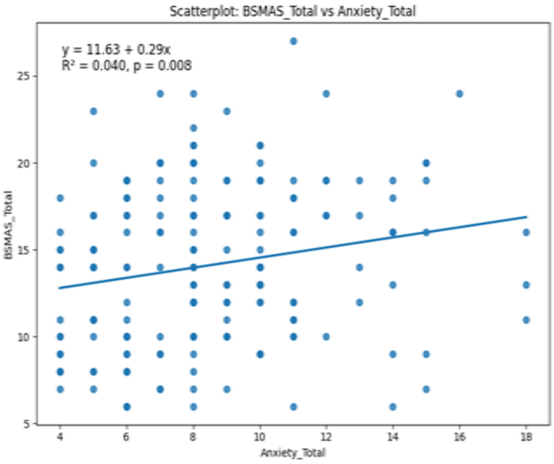
5.2.1 Scatterplot



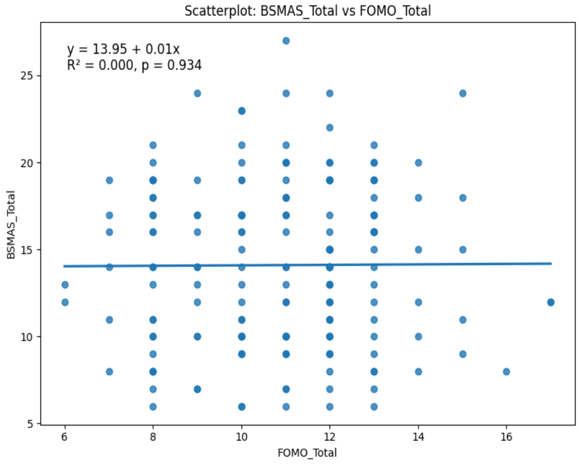
BSMAS_Total vs Self Esteem_Total



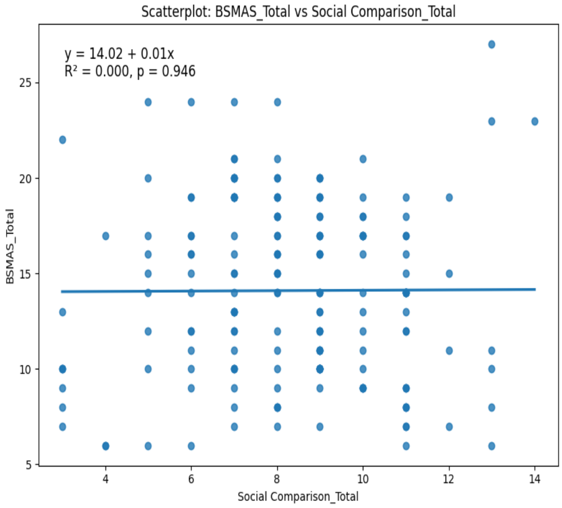
BSMAS_Total vs Depression_Total



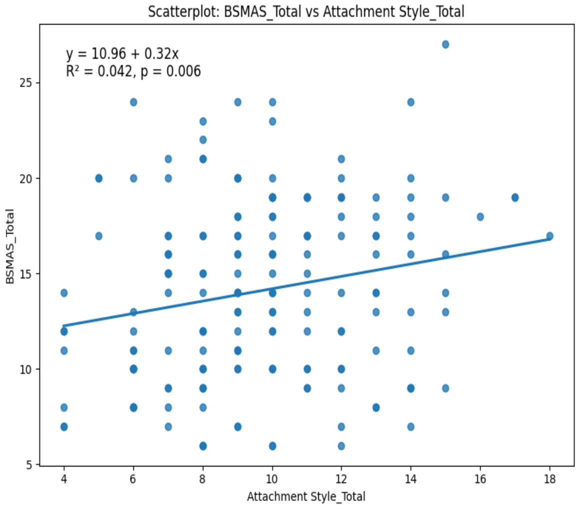
BSMAS_Total vs Anxiety_Total



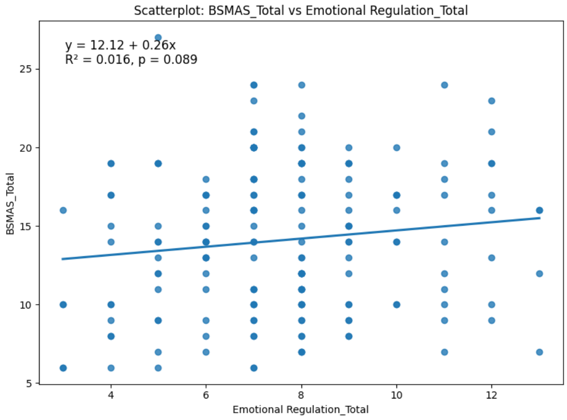
BSMAS_Total vs FOMO_Total



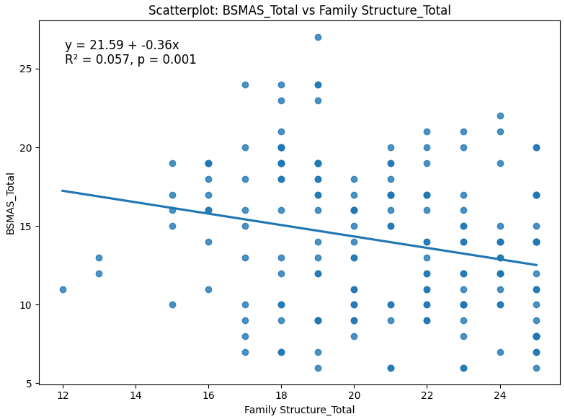
BSMAS_Total vs Social Comparison_Total



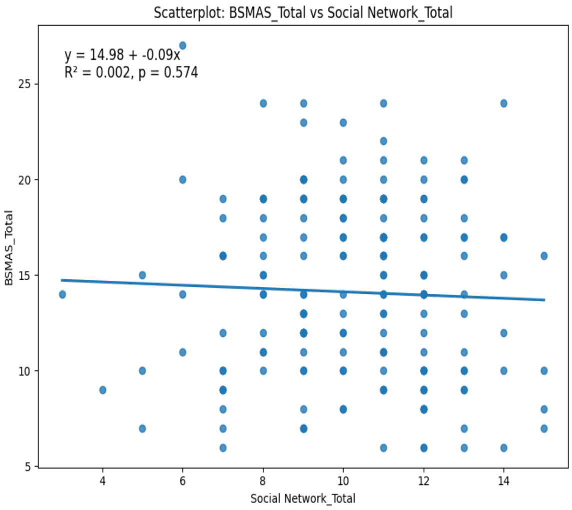
BSMAS_Total vs Emotional Regulation_Total



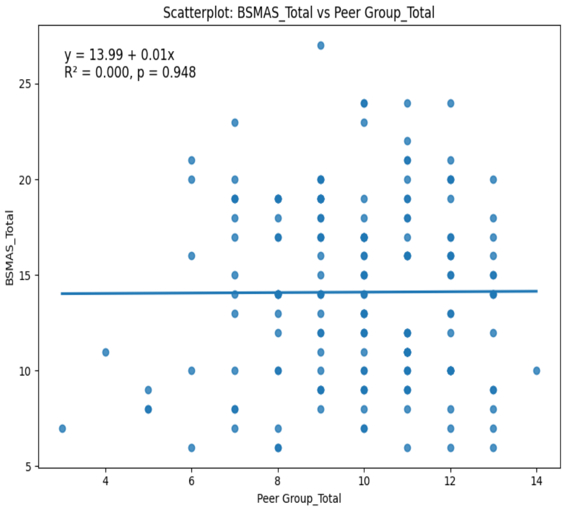
BSMAS_Total vs Attachment Style_Total



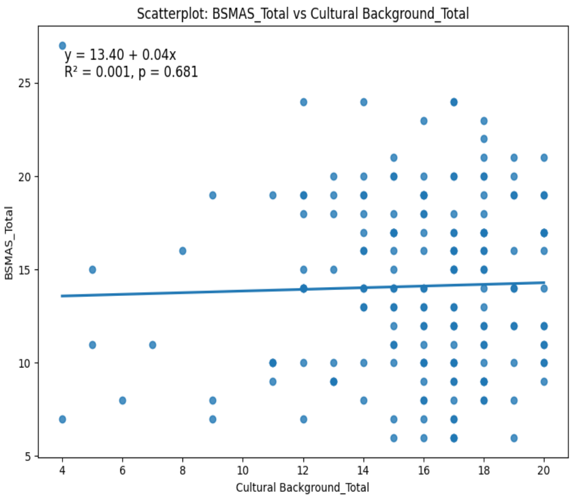
BSMAS_Total vs Family Structure_Total



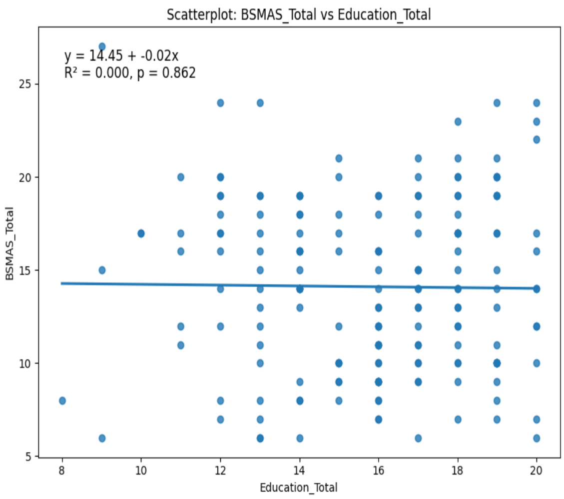
BSMAS_Total vs Social Network_Total



BSMAS_Total vs Peer Group_Total



BSMAS_Total vs Cultural Background_Total



BSMAS_Total vs Education_Total

Fig. 2. Scatter Plots of BSMAS_Total vs Each of Psychological and Social Factor

From Fig. 2, There is no any proper pattern seen to indicate positive or negative correlation. Hence, we have to check correlation using correlation tests. There is no strong linear correlation among the BSMAS and other psychological and social factors.

5.2.2 Correlation test

To check the relations between factors we have used two approaches first is factor wise correlation with social media addiction and next is question wise correlation with social media addiction. Following tables shows the Pearson, spearman and Kendall correlations. Also we use chi-square table of contingency for checking association between factors and social media addiction.

	Pearson		Spearman		Kendall	
	coefficient	**p value**	**coefficient**	**p value**	**coefficient**	**p value**
S_E	0.158	0.035	0.157	0.035	0.117	0.031
Dep	0.296	0.000	0.282	0.000	0.206	0.000
Anx	0.199	0.008	0.221	0.003	0.157	0.004
FOMO	0.006	0.934	0.012	0.877	0.01	0.858
S_C	0.005	0.946	-0.057	0.447	-0.05	0.365
E_R	0.128	0.089	0.112	0.136	0.082	0.133
A_S	0.205	0.006	0.195	0.009	0.143	0.008
F_S	-0.239	0.001	-0.255	0.001	-0.184	0.001
S_N	-0.042	0.574	-0.056	0.458	-0.041	0.45
P_G	0.005	0.948	-0.039	0.603	-0.038	0.493
C_B	0.031	0.681	0.032	0.674	0.021	0.704
Edu	-0.013	0.862	0.007	0.925	0.008	0.877

Table 3: Factor wise Correlation test with BSMAS_Total

Dep (Depression) shows moderate positive correlation across methods (Pearson: 0.296, Spearman: 0.282, Kendall: 0.206), with $p < 0.001$. Suggests depression is moderately related to BSMAS_Total. **Anx (Anxiety)**: Moderate positive correlation (Pearson: 0.199, Spearman: 0.221, Kendall: 0.157), **S_E (Self Esteem)**: Moderate positive correlation (Pearson: 0.158, Spearman: 0.157, Kendall: 0.117), **A_S (Attachment Style)**: moderate positive correlation ((Pearson: 0.205, Spearman: 0.195, Kendall: 0.143), all highly significant. **F_S (Family Structure)**: Strong negative correlation (Pearson: -0.239, Spearman: -0.255, Kendall: -0.184), significant in all methods.

P_G (Peer Group), **S_C (Social Comparison)** and **FOMO (FOMO)**: Coefficients are close to zero, and p-values indicate no significant correlation with BSMAS_Total.

	Pearson		Spearman		Kendall	
	coefficient	**p value**	**coefficient**	**p value**	**coefficient**	**p value**
SE_1	0.031	0.676	0.066	0.38	0.05	0.388
SE_2	-0.048	0.526	-0.028	0.712	-0.023	0.696
SE_3	0.215	0.004	0.235	0.002	0.189	0.001
SE_4	0.278	0.000	0.263	0.000	0.211	0.000
Dep_1	0.176	0.019	0.165	0.027	0.128	0.026
Dep_2	0.066	0.377	0.063	0.4	0.051	0.381
Dep_3	0.374	0.000	0.378	0.000	0.29	0.000
Dep_4	0.166	0.027	0.161	0.031	0.132	0.026
Dep_5	0.165	0.028	0.126	0.093	0.103	0.088
Anx_1	0.08	0.287	0.082	0.273	0.063	0.284
Anx_2	0.177	0.018	0.179	0.017	0.139	0.016
Anx_3	0.193	0.01	0.21	0.005	0.168	0.005

Anx_4	0.111	0.139	0.131	0.082	0.1	0.085
FOMO_1	0.19	0.011	0.183	0.014	0.147	0.014
FOMO_2	-0.253	0.001	-0.271	0.000	-0.208	0.000
FOMO_3	0.157	0.036	0.164	0.028	0.127	0.027
FOMO_4	-0.01	0.897	-0.018	0.815	-0.012	0.831
SC_1	0.292	0.000	0.282	0.000	0.218	0.000
SC_2	-0.091	0.227	-0.088	0.243	-0.068	0.231
SC_3	-0.152	0.042	-0.153	0.041	-0.123	0.031
ER_1	0.069	0.358	0.09	0.23	0.068	0.234
ER_2	0.092	0.222	0.091	0.228	0.07	0.221
ER_3	0.078	0.298	0.082	0.278	0.065	0.261
AS_1	0.066	0.379	0.055	0.462	0.042	0.457
AS_2	0.059	0.434	0.08	0.29	0.062	0.278
AS_3	0.318	0.000	0.328	0.000	0.254	0.000
AS_4	0.065	0.387	0.092	0.22	0.069	0.23
FS_1	-0.044	0.561	-0.105	0.162	-0.084	0.169
FS_2	-0.245	0.001	-0.244	0.001	-0.19	0.001
FS_3	0.045	0.553	0.007	0.921	0.008	0.902
FS_4	-0.084	0.265	-0.155	0.039	-0.124	0.036
FS_5	-0.218	0.003	-0.236	0.001	-0.176	0.002
SN_1	-0.132	0.079	-0.137	0.067	-0.107	0.066
SN_2	-0.089	0.238	-0.076	0.315	-0.06	0.3
SN_3	0.109	0.147	0.108	0.151	0.086	0.132
PG_1	0.024	0.751	0.024	0.754	0.019	0.742
PG_2	0.061	0.414	0.033	0.665	0.021	0.721
PG_3	-0.08	0.289	-0.128	0.087	-0.103	0.075
CB_1	0	0.997	-0.024	0.746	-0.02	0.733
CB_2	0.098	0.193	0.102	0.175	0.079	0.167
CB_3	0.033	0.663	0.023	0.761	0.017	0.774
CB_4	-0.042	0.573	-0.048	0.524	-0.038	0.51
Edu_1	-0.053	0.483	-0.028	0.706	-0.023	0.698
Edu_2	-0.012	0.868	-0.029	0.704	-0.024	0.691
Edu_3	-0.148	0.049	-0.151	0.043	-0.118	0.041
Edu_4	0.143	0.055	0.115	0.124	0.087	0.129

Table 4: Question wise Correlation test with BSMAS_Total

The psychological factor Self-esteem Question number 3,4; Depression Question number 1,3,4,5; Anxiety Question number 2,3; FOMO Question number 1,2,3; Social comparison Question number 1,3; Attachment style Question number 3; shows significant correlations with social media addiction (BSMAS)

The social factor Family Structure Question number 2,5; and Education Question number shows significant correlation with BSMAS

5.2.3 Chi-square contingency table

		BSMAS					Cramer's V square	Kappa
		High	**Low**	**Moderate**	**Very High**	**All**		
Self Esteem	high	3(1.676)	4(2.235)	7(3.911)	1(0.559)	15(8.38)	0.025	0.052
	low	17(9.497)	36(20.112)	34(18.994)	0(0)	87(48.603)		
	Moderate	21(11.732)	21(11.732)	31(17.318)	4(2.235)	77(43.017)		
	very high	41(22.905)	61(34.078)	72(40.223)	5(2.793)	179(100)		
	All	3(1.676)	2(1.117)	5(2.793)	2(1.117)	12(6.704)		
Depression	high	17(9.497)	26(14.525)	35(19.553)	1(0.559)	79(44.134)	0.043	0.003
	low	14(7.821)	11(6.145)	20(11.173)	2(1.117)	47(26.257)		
	Moderate	0(0)	0(0)	1(0.559)	0(0)	1(0.559)		
	very high	7(3.911)	22(12.291)	11(6.145)	0(0)	40(22.346)		
	very low	41(22.905)	61(34.078)	72(40.223)	5(2.793)	179(100)		
	All	8(4.469)	5(2.793)	9(5.028)	1(0.559)	23(12.849)		
Anxiety	high	13(7.263)	35(19.553)	23(12.849)	1(0.559)	72(40.223)	0.039	0.17
	low	20(11.173)	20(11.173)	38(21.229)	2(1.117)	80(44.693)		
	Moderate	0(0)	1(0.559)	2(1.117)	1(0.559)	4(2.235)		
	very high	41(22.905)	61(34.078)	72(40.223)	5(2.793)	179(100)		
FOMO	All	17(9.497)	23(12.849)	32(17.877)	2(1.117)	74(41.341)		
	high	1(0.559)	2(1.117)	5(2.793)	0(0)	8(4.469)	0.008	-0.046
	low	23(12.849)	35(19.553)	33(18.436)	3(1.676)	94(52.514)		
	Moderate	0(0)	1(0.559)	2(1.117)	0(0)	3(1.676)		
	very high	41(22.905)	61(34.078)	72(40.223)	5(2.793)	179(100)		
	All	5(2.793)	15(8.38)	13(7.263)	1(0.559)	34(18.994)		
Social Comparison	high	4(2.235)	14(7.821)	15(8.38)	2(1.117)	35(19.553)	0.024	-0.008
	low	32(17.877)	32(17.877)	44(24.581)	2(1.117)	110(61.45)		
	Moderate	41(22.905)	61(34.078)	72(40.223)	5(2.793)	179(100)		
	All	6(3.352)	7(3.911)	9(5.028)	1(0.559)	23(12.849)		
Emotional Regulations	high	6(3.352)	19(10.615)	24(13.408)	1(0.559)	50(27.933)	0.015	-0.008
	low	29(16.201)	35(19.553)	39(21.788)	3(1.676)	106(59.21)		
	Moderate	41(22.905)	61(34.078)	72(40.223)	5(2.793)	179(100)		
	All	12(6.704)	14(7.821)	16(8.939)	2(1.117)	44(24.581)		
Attachment Style	high	5(2.793)	20(11.173)	16(8.939)	1(0.559)	42(23.464)	0.024	0.09
	low	21(11.732)	27(15.084)	39(21.788)	2(1.117)	89(49.721)		
	Moderate	3(1.676)	0(0)	1(0.559)	0(0)	4(2.235)		
	very high	41(22.905)	61(34.078)	72(40.223)	5(2.793)	179(100)		
	All	22(12.291)	21(11.732)	21(11.732)	5(2.793)	69(38.547)		
Family Structure	high	0(0)	1(0.559)	0(0)	0(0)	1(0.559)	0.041	0.044
	low	5(2.793)	2(1.117)	9(5.028)	0(0)	16(8.939)		
	Moderate	14(7.821)	37(20.67)	42(23.464)	0(0)	93(51.955)		
	very high	41(22.905)	61(34.078)	72(40.223)	5(2.793)	179(100)		
	All	17(9.497)	34(18.994)	37(20.67)	2(1.117)	90(50.279)		
Social Network	high	1(0.559)	4(2.235)	3(1.676)	1(0.559)	9(5.028)	0.017	-0.023
	low	23(12.849)	23(12.849)	32(17.877)	2(1.117)	80(44.693)		
	Moderate	41(22.905)	61(34.078)	72(40.223)	5(2.793)	179(100)		
	All	15(8.38)	30(16.76)	27(15.084)	2(1.117)	74(41.341)		

Peer Groups	high	2(1.117)	7(3.911)	1(0.559)	0(0)	10(5.587)	0.031	0.062
	low	24(13.408)	24(13.408)	44(24.581)	3(1.676)	95(53.073)		
	Moderate	41(22.905)	61(34.078)	72(40.223)	5(2.793)	179(100)		
	All	14(7.821)	13(7.263)	24(13.408)	2(1.117)	53(29.609)		
Cultural Background	high	0(0)	4(2.235)	1(0.559)	1(0.559)	6(3.352)	0.029	0.001
	low	2(1.117)	6(3.352)	1(0.559)	0(0)	9(5.028)		
	Moderate	25(13.966)	38(21.229)	46(25.698)	2(1.117)	111(62.01)		
	very high	41(22.905)	61(34.078)	72(40.223)	5(2.793)	179(100)		
	All	16(8.939)	18(10.056)	24(13.408)	2(1.117)	60(33.52)		
Education	high	1(0.559)	3(1.676)	6(3.352)	1(0.559)	11(6.145)	0.013	0.018
	low	24(13.408)	40(22.346)	42(23.464)	2(1.117)	108(60.33)		
	Moderate	41(22.905)	61(34.078)	72(40.223)	5(2.793)	179(100)		
	very high	29(16.201)	29(16.201)	30(16.76)	4(2.235)	92(51.397)		
	All	44(24.581)	52(29.05)	64(35.754)	8(4.469)	168(93.85)		

Table 5: Chi-square contingency table of BSMAS and psychological, social factors

The analysis of the contingency table indicates no significant association between the BSMAS scores and the psychological and social factors, based on the general criteria for association.

- **Cramer's V square**: A value greater than 0.40 is typically required to indicate a moderate association.
- **Kappa coefficient**: A value greater than 0.60 is considered indicative of a moderate agreement.

However, in this case, these thresholds are not met, suggesting no meaningful relationship between the variables.

Additionally, due to the small dataset, certain assumptions of the chi-square test for association have been violated. To address this, it is recommended to reduce the number of categories to ensure more robust results.

		BSMAS			Chi_sq_p	Fishers_p	Cramer's V Square	Kappa
		high	**low**	**All**				
Self Esteem	high	4(2.235)	34(18.994)	38(21.229)	0.081	0.097	0.017	0.097
	low	5(2.793)	136(75.978)	141(78.771)				
	All	9(5.028)	170(94.972)	179(100)				
Depression	high	6(3.352)	32(17.877)	38(21.229)	0.243	0.245	0.008	0.08
	low	13(7.263)	128(71.508)	141(78.771)				
	All	19(10.615)	160(89.385)	179(100)				
Anxiety	high	6(3.352)	32(17.877)	38(21.229)	0.381	0.398	0.004	0.062
	low	15(8.38)	126(70.391)	141(78.771)				
	All	21(11.732)	158(88.268)	179(100)				
FOMO	high	8(4.469)	30(16.76)	38(21.229)	0.902	1	0	-0.009
	low	31(17.318)	110(61.453)	141(78.771)				
	All	39(21.788)	140(78.212)	179(100)				
Social Comparison	high	7(3.911)	31(17.318)	38(21.229)	0.104	0.113	0.015	-0.119
	low	45(25.14)	96(53.631)	141(78.771)				
	All	52(29.05)	127(70.95)	179(100)				

Emotional Regulations	high	7(3.911)	31(17.318)	38(21.229)	0.921	1	0	0.007
	low	25(13.966)	116(64.804)	141(78.771)				
	All	32(17.877)	147(82.123)	179(100)				
Attachment Style	high	8(4.469)	30(16.76)	38(21.229)	0.639	0.641	0.001	0.035
	low	25(13.966)	116(64.804)	141(78.771)				
	All	33(18.436)	146(81.564)	179(100)				
Family Structure	high	37(20.67)	1(0.559)	38(21.229)	0.537	1	0.002	0.01
	low	134(74.86)	7(3.911)	141(78.771)				
	All	171(95.531)	8(4.469)	179(100)				
Social Network	high	22(12.291)	16(8.939)	38(21.229)	0.556	0.576	0.002	-0.031
	low	89(49.721)	52(29.05)	141(78.771)				
	All	111(62.011)	68(37.989)	179(100)				
Peer Groups	high	18(10.056)	20(11.173)	38(21.229)	0.066	0.092	0.019	-0.098
	low	90(50.279)	51(28.492)	141(78.771)				
	All	108(60.335)	71(39.665)	179(100)				
Cultural Background	high	32(17.877)	6(3.352)	38(21.229)	0.803	0.798	0	-0.008
	low	121(67.598)	20(11.173)	141(78.771)				
	All	153(85.475)	26(14.525)	179(100)				
Education	high	31(17.318)	7(3.911)	38(21.229)	0.307	0.295	0.006	-0.03
	low	124(69.274)	17(9.497)	141(78.771)				
	All	155(86.592)	24(13.408)	179(100)				

Table 6: Chi-square contingency table (2x2) of BSMAS and psychological, social factors

- **Effect Size (Cramer's V square):** Across the table, most Cramer's V square values are very small (near 0), indicating weak or negligible associations. The strongest effect size is observed for **Self Esteem (Cramer's V square = 0.017)**, though it is still below the threshold for moderate association.
- **Agreement (Kappa):** The Kappa values across the table are mostly very low, showing minimal agreement. The highest Kappa is observed for **Self Esteem (Kappa = 0.097),** suggesting some agreement but still below the threshold for moderate levels.
- **Significant Factors:** No any factor show significant associations with BSMAS scores.
- **Implications:** The absence of associations may be due to small sample size or underlying complexity in the relationship between BSMAS scores and these factors. Further research with larger datasets and refined methods is recommended.

5.2.4 Multiple linear regression

This approach is used to check the relation with regression coefficient with the help of model outputs R^2, coefficients and p-value of F-statistics and t-statistics resp. Here we used two approaches first is regress single factor with social media addiction factor and another one is regress all factors at a time with social media addiction. In both approaches social media addiction score is dependent variable.

	Independent Variable	Coeff. (B)	SE	t-statistic	p value	R^2	F-statistic	p (F-statistic)
Psychological Factors	Self Esteem	7.689	2.614	2.942	**0.004**	0.025	4.529	**0.035**
	Depression	0.094	0.139	0.674	0.501	0.088	17.013	0.000
	Anxiety	0.325	0.115	2.822	**0.005**	0.040	7.281	**0.008**
	FOMO	0.023	0.127	0.183	0.855	0.000	0.007	0.934
	Social Comparison	-0.064	0.161	-0.399	0.691	0.000	0.005	0.946
	Emotional Regulation	-0.099	0.148	-0.668	0.505	0.016	2.927	0.089
	Attachment Style	0.181	0.154	1.178	0.241	0.042	7.734	**0.006**
Social Factors	Family Structure	0.211	0.124	1.708	0.089	0.057	10.699	**0.001**
	Social Network	-0.086	0.152	-0.563	0.574	0.002	0.317	0.574
	Peer Groups	0.011	0.170	0.066	0.948	0.000	0.004	0.948
	Cultural Background	0.045	0.108	0.412	0.681	0.001	0.170	0.681
	Education	-0.022	0.127	-0.174	0.862	0.000	0.030	0.862

Table 7: Regression with single factor total as independent variable

	Independent Variable	Coefficient (B)	Standard Error (SE)	t-statistic	p value	R^2	F-statistic	p (F-statistic)
Psychological Factor	Self Esteem	0.285	0.134	2.128	**0.035**	0.1219	3.3918	**0.0020**
	Depression	0.394	0.096	4.125	**0.000**			
	Anxiety	0.291	0.108	2.698	**0.008**			
	FOMO	0.014	0.163	0.083	0.934			
	Social Comparison	0.010	0.150	0.067	0.946			
	Emotional Regulation	0.260	0.152	1.711	**0.089**			
	Attachment Style	0.324	0.117	2.781	**0.006**			
Social Factors	Family Structure	-0.383	0.114	-3.364	**0.001**	0.0657	2.4321	**0.0369**
	Social Network	-0.089	0.164	-0.545	0.587			

	Peer Groups	0.033	0.182	0.181	0.856			
	Cultural Background	0.133	0.116	1.150	0.252			
	Education	-0.081	0.128	-0.631	0.529			

Table 8: Regression with all psychological and social factors as independent variables.

- **Independent Variable Analysis:**

  When analyzed individually, variables such as Self-Esteem, Anxiety, Attachment Style, and Family Structure satisfy the p-value criteria for statistical significance, indicating some relationship with social media addiction (BSMAS score).

  However, when analyzed together using a multivariate model, these variables also satisfy the p-value threshold, but the coefficients and the R^2 value remain very low. This indicates a weak overall relationship between psychological and social factors and social media addiction.

- **Bivariate Analysis:**

  From the bivariate analysis, it is evident that the association between psychological and social factors (independent variables) and social media addiction (dependent variable) is weak. This suggests no strong or consistent linear relationship among these variables.

- **Addressing the Problem with Logistic Regression:**

  Due to the lack of a clear linear relationship, a logistic regression approach was applied to model the relationship between the independent variables and the binary outcome of social media addiction (addicted vs. not addicted).

  One key challenge in this approach was determining how to categorize individuals into "addicted" and "not addicted" groups based on the BSMAS scores.

- ***Empirical Cumulative Frequency* Function (ECDF) Plot Method:**

  To address the challenge of categorizing individuals, the ECDF plot method was used. In this method, the "splitting point" between the categories of addicted and not addicted was determined by identifying the point on the ECDF curve where the graph demonstrates a significant change in behavior. This point indicates the threshold at which individuals' behavior shifts, providing a data-driven basis for categorization.

  The use of the ECDF plot method provides an objective and empirical way to classify individuals into addiction categories, ensuring consistency and reducing subjectivity in determining the threshold.

.

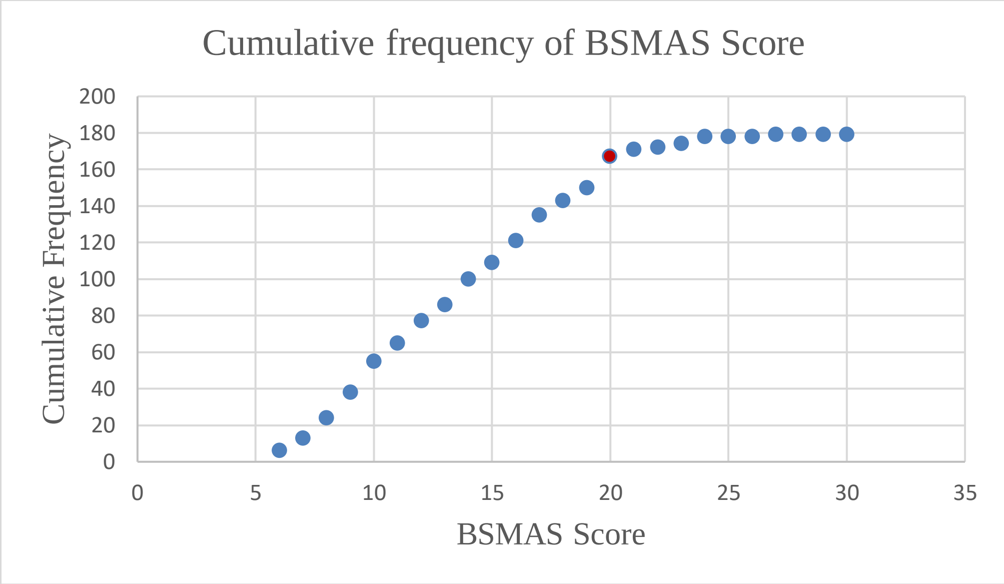


Fig. 3: Empirical cdf of BSMAS Score

In above Fig. 3., the highlighted point (20) shows cut of the graph means from above that point is all points are considered as addicted to Social Media. This analysis we used for predictive analysis.

5.3 Statistical models

5.3.1 Binomial logistic regression factor wise

```
Name: Y, Length: 179, dtype: int64: []
                             OrderedModel Results
==============================================================================
Dep. Variable:                      Y   Log-Likelihood:                -63.572
Model:                   OrderedModel   AIC:                             153.1
Method:            Maximum Likelihood   BIC:                             194.6
Date:                Fri, 22 Nov 2024
Time:                        16:17:40
No. Observations:                 179
Df Residuals:                     166
Df Model:                          12
===============================================================================================
                               coef    std err          z      P>|z|      [0.025      0.975]
-----------------------------------------------------------------------------------------------
Self Esteem_Total           -0.0500      0.105     -0.478      0.632      -0.255       0.155
Depression_Total             0.0641      0.083      0.772      0.440      -0.099       0.227
Anxiety_Total                0.0605      0.086      0.706      0.480      -0.108       0.228
FOMO_Total                   0.0438      0.111      0.396      0.692      -0.173       0.260
Social Comparison_Total     -0.0547      0.111     -0.495      0.621      -0.272       0.162
Emotional Regulation_Total   0.0764      0.113      0.675      0.500      -0.145       0.298
Attachment Style_Total      -0.1328      0.098     -1.348      0.178      -0.326       0.060
Family Structure_Total      -0.0689      0.094     -0.729      0.466      -0.254       0.116
Social Network_Total        -0.0281      0.116     -0.243      0.808      -0.255       0.199
Peer Group_Total            -0.0112      0.125     -0.090      0.929      -0.257       0.235
Cultural Background_Total    0.0093      0.080      0.116      0.907      -0.147       0.166
Education_Total              0.0738      0.091      0.816      0.415      -0.104       0.251
0/1                          1.7390      4.169      0.417      0.677      -6.432       9.910
===============================================================================================
```

Fig 4: Output of Binomial Logistic Regression

All predictors in the model do not show a statistically significant relationship with the outcome. This insignificance might be driven by the aggregated factor sum, potentially masking the contribution of individual items within the factor. To address this, instead of using the summed factor score, we include all individual items (questions) as independent variables in the model for a more detailed analysis.

5.3.2 Binomial logistic regression Question wise

In this analysis, all individual questions are used as independent variables, with the categorized BSMAS serving as the dependent variable. To refine the model, the backward elimination method is employed to systematically remove questions that do not contribute significantly to the model. The results of the Binomial Logistic Regression analysis, performed on a question-by-question basis, are presented in Figure 5.

```
Results for Y (Final Model with Significant Variables):

                           Logit Regression Results
==============================================================================
Dep. Variable:                      Y   No. Observations:                  179
Model:                          Logit   Df Residuals:                      177
Method:                           MLE   Df Model:                            1
Date:                Fri, 22 Nov 2024   Pseudo R-squ.:                 0.01530
Time:                        16:19:40   Log-Likelihood:                -65.688
converged:                       True   LL-Null:                       -66.709
Covariance Type:            nonrobust   LLR p-value:                    0.1531
================================================================================
                   coef    std err          z      P>|z|      [0.025      0.975]
--------------------------------------------------------------------------------
const           -3.2755      1.013     -3.232      0.001      -5.262      -1.289
Peer Group_3     0.3295      0.241      1.369      0.171      -0.142       0.801
================================================================================
Performance Metrics for Y:

Accuracy: 0.8771
F1 Score: 0.0000
Precision: 0.0000
Recall: 0.0000
Confusion Matrix:
[[157   0]
 [ 22   0]]
```

Fig 5: Output of Binomial Logistic Regression Question wise

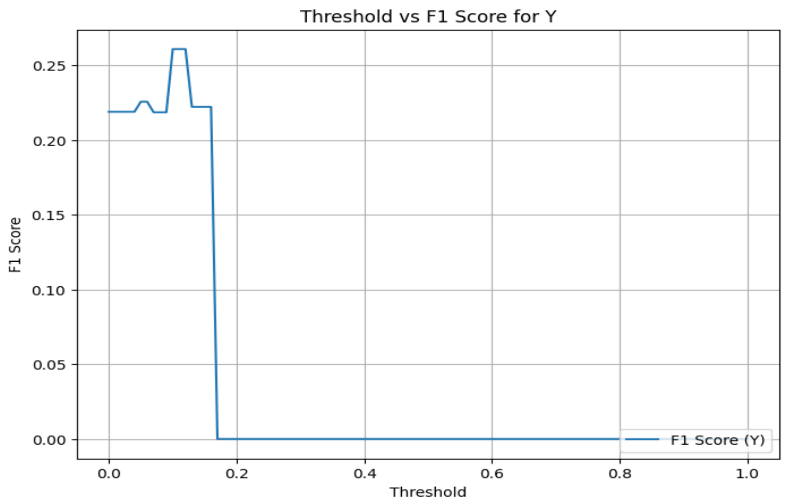


Fig 6: F1 Score Plot

Peer Group_3 positively impact the likelihood of Y (Categorized BSMAS).

The results presented in Figure 5, indicate that the model is statistically significant. However, a class imbalance problem was observed in the data, which could affect the model's performance. To address this issue, we applied the **SMOTE (*Synthetic Minority Oversampling Technique*)** to balance the classes, ensuring a more robust and reliable analysis.

5.3.3 Binomial logistic regression using SMOTE

```
Results for Y (Final Model with Significant Variables):

                           Logit Regression Results
==============================================================================
Dep. Variable:                      Y   No. Observations:                  314
Model:                          Logit   Df Residuals:                      297
Method:                           MLE   Df Model:                           16
Date:                Fri, 22 Nov 2024   Pseudo R-squ.:                  0.5967
Time:                        16:21:41   Log-Likelihood:                -87.768
converged:                      False   LL-Null:                       -217.65
Covariance Type:            nonrobust   LLR p-value:                 5.130e-46
======================================================================================
                         coef    std err          z      P>|z|      [0.025      0.975]
--------------------------------------------------------------------------------------
const                  1.5711      3.062      0.513      0.608      -4.431       7.573
Self Esteem_2         -1.4376      0.335     -4.295      0.000      -2.094      -0.782
Self Esteem_4          1.3994      0.417      3.357      0.001       0.582       2.216
Depression_3           1.0359      0.243      4.270      0.000       0.560       1.511
Depression_4          -1.2753      0.297     -4.299      0.000      -1.857      -0.694
Depression_5          -1.0813      0.363     -2.978      0.003      -1.793      -0.370
FOMO_2                -1.2517      0.215     -5.820      0.000      -1.673      -0.830
FOMO_3                 1.0061      0.236      4.255      0.000       0.543       1.469
Attachment Style_4    -1.4189      0.334     -4.244      0.000      -2.074      -0.764
Family Structure_2    -1.4741      0.223     -6.616      0.000      -1.911      -1.037
Family Structure_3     1.6361      0.581      2.815      0.005       0.497       2.775
Social Network_2      -1.5088      0.276     -5.457      0.000      -2.051      -0.967
Social Network_3       0.8102      0.204      3.969      0.000       0.410       1.210
Peer Group_2          -0.7690      0.236     -3.265      0.001      -1.231      -0.307
Peer Group_3           0.9899      0.303      3.262      0.001       0.395       1.585
Education_1            1.0682      0.361      2.957      0.003       0.360       1.776
Education_3           -0.7690      0.195     -3.941      0.000      -1.151      -0.387
======================================================================================
Performance Metrics for Y:

Accuracy: 0.8917
F1 Score: 0.8938
Precision: 0.8773
Recall: 0.9108
Confusion Matrix:
[[137  20]
 [ 14 143]]
```

Fig. 7: Binomial logistic regression using SMOTE

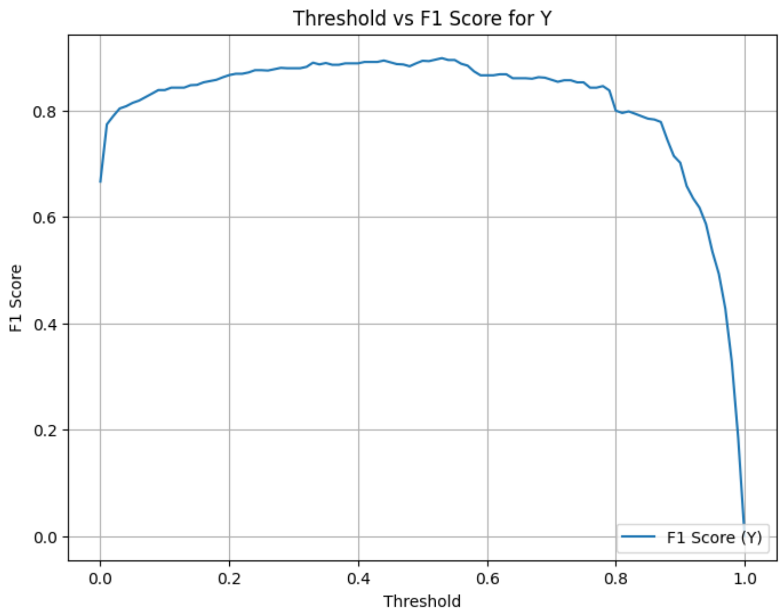


Fig. 8: F1 Score Plot

General Model Overview

This logistic regression model predicts the probability of Y (e.g., high BSMAS scores or presence of social media addiction) based on a set of independent variables. It uses *Maximum Likelihood Estimation* (MLE) for parameter estimation. The model demonstrates good fit, as indicated by a Pseudo R-squared of 0.5967, which explains about 59.67% of the variance in the dependent variable. The Log-Likelihood (-87.768) indicates the likelihood of observing the data given the model parameters.

Model Performance Metrics

- Accuracy: 89.17% – Proportion of correctly predicted outcomes.
- F1 Score: 89.38% – Balance between precision and recall.
- Precision: 87.73% – Proportion of true positive predictions to all positive predictions.
- Recall: 91.08% – Proportion of actual positives correctly predicted.
- Confusion Matrix:
    - True Positives (TP): 137
    - True Negatives (TN): 143
    - False Positives (FP): 20
    - False Negatives (FN): 14

The metrics indicate the model is highly accurate and well-calibrated for predicting Y.

Variable	Coeff.	Odds Ratio	p value	Interpretation
Self Esteem_2	-1.4376	0.238	0.000	A one-unit increase in Self Esteem_2 decreases the odds of higher BSMAS scores by 76.2%, suggesting lower self-esteem significantly reduces the risk of addiction.
Self Esteem_4	1.3994	4.051	0.001	A one-unit increase in Self Esteem_4 increases the odds of higher BSMAS scores by 305.1%, indicating elevated self-esteem may drive social media usage for validation.
Depression_3	1.0359	2.817	0.000	A one-unit increase in Depression_3 increases the odds of higher BSMAS scores by 181.7%, suggesting moderate depression is closely linked to excessive social media use.
Depression_4	-1.2753	0.279	0.000	A one-unit increase in Depression_4 decreases the odds of higher BSMAS scores by 72.1%, showing severe depression can reduce the dependence on social media.
Depression_5	-1.0813	0.339	0.003	A one-unit increase in Depression_5 decreases the odds of higher BSMAS scores by 66.1%, likely reflecting emotional withdrawal.
FOMO_2	-1.2517	0.286	0.000	A one-unit increase in FOMO_2 decreases the odds of higher BSMAS scores by 71.4%, suggesting low fear of missing out reduces addiction risk.
FOMO_3	1.0061	2.735	0.000	A one-unit increase in FOMO_3 increases the odds of higher BSMAS scores by 173.5%, reflecting social pressure and engagement.
Attachment Style_4	-1.4189	0.242	0.000	A one-unit increase in Attachment Style_4 decreases the odds of higher BSMAS scores by 75.8%, indicating avoidant attachment reduces addiction tendencies.
Family Structure_2	-1.4741	0.229	0.000	A one-unit increase in Family Structure_2 decreases the odds of higher BSMAS scores by 77.1%, showing strong family connections lower addiction risk.
Family Structure_3	1.6361	5.134	0.005	A one-unit increase in Family Structure_3 increases the odds of higher BSMAS scores by 413.4%, indicating poor family dynamics can increase social media addiction.
Social Network_2	-1.5888	0.204	0.000	A one-unit increase in Social Network_2 decreases the odds of higher BSMAS scores by 79.6%, highlighting offline networks as a protective factor.
Social Network_3	0.8102	2.248	0.000	A one-unit increase in Social Network_3 increases the odds of higher BSMAS scores by 124.8%, suggesting online networks drive excessive use.
Peer Group_2	-0.7690	0.464	0.001	A one-unit increase in Peer Group_2 decreases the odds of higher BSMAS scores by 53.6%, showing supportive peer groups mitigate addiction.

Peer Group_3	1.9399	6.955	0.001	A one-unit increase in Peer Group_3 increases the odds of higher BSMAS scores by 595.5%, indicating unhealthy peer influences elevate addiction risks.
Education_1	1.0682	2.911	0.003	A one-unit increase in Education_1 increases the odds of higher BSMAS scores by 191.1%, showing a positive link between academic pressure and social media use.
Education_3	-0.7690	0.464	0.000	A one-unit increase in Education_3 decreases the odds of higher BSMAS scores by 53.6%, indicating reduced dependence among students with balanced educational practices.

Table 9: Logistic Regression Results with Interpretation

5.4 Machine learning algorithms

For further analysis, we employed machine learning algorithms such as Random Forest and XGBoost to develop comparative models and evaluate their performance against the statistical Binomial Logistic Regression model.

5.4.1 Random Forest and XGBoost

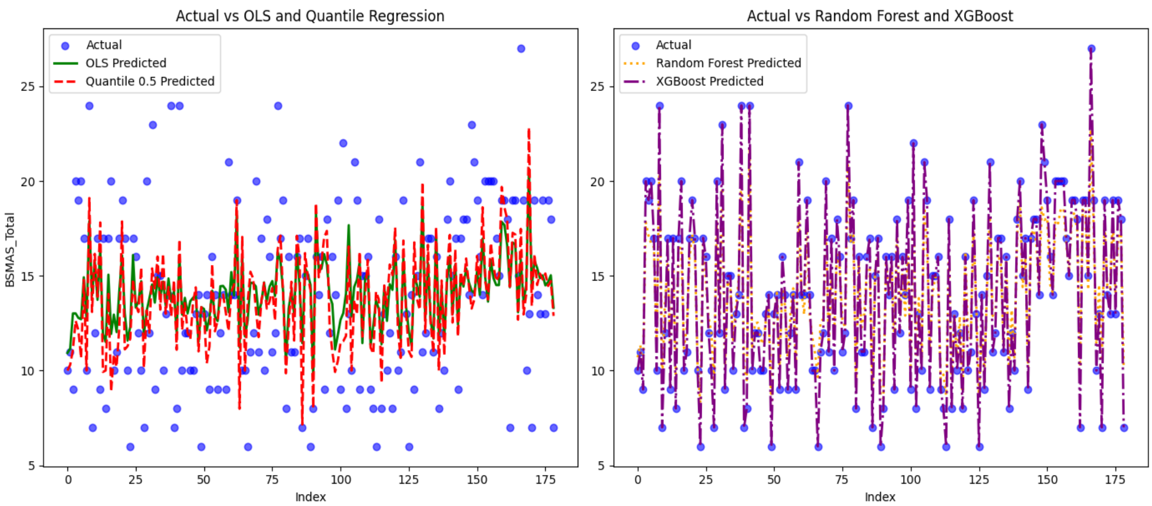


Fig. 9: Statistical Models vs Machin learning Models

Sr. No.	Model	RMSE
1	OLS	4.2982
2	Quantile Regression (0.5)	4.4493
3	Random Forest	1.6817
4	XGBoost	0.0011

Table 10: Model Performance Metrics (RMSE):

XGBoost outperforms all other models with the lowest RMSE, followed by Random Forest, while OLS and Quantile Regression show comparatively poor predictive performance. This suggests that advanced machine learning models like XGBoost and Random Forest are better suited for this nonlinear dataset.

5.4.2 Partial Dependency plots

5.4.2.1 Partial Dependency Plots for Random Forest

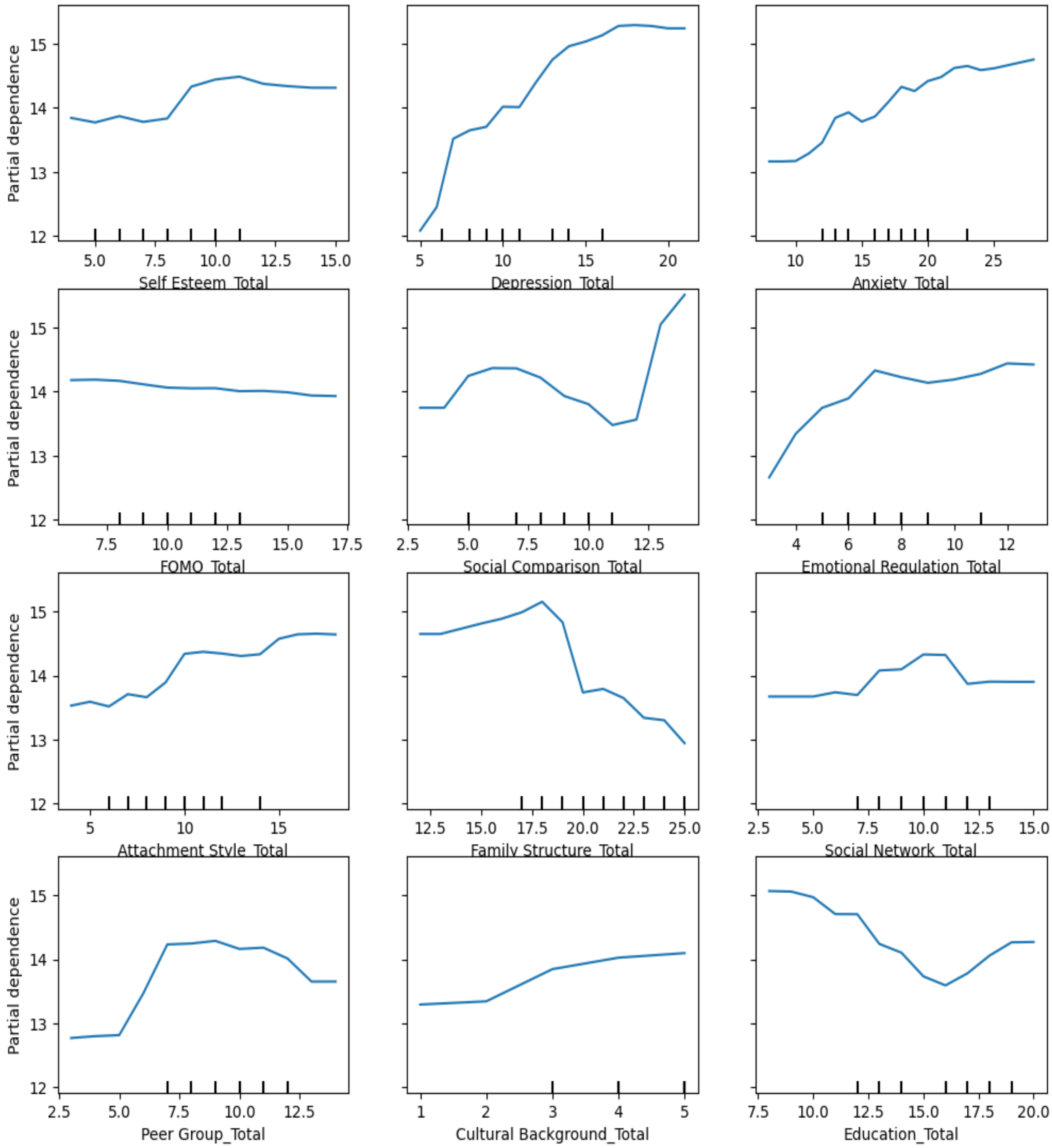


Fig 10: Partial Dependency Plots for Random Forest

5.4.2.2 Partial Dependency Plots for XGboost

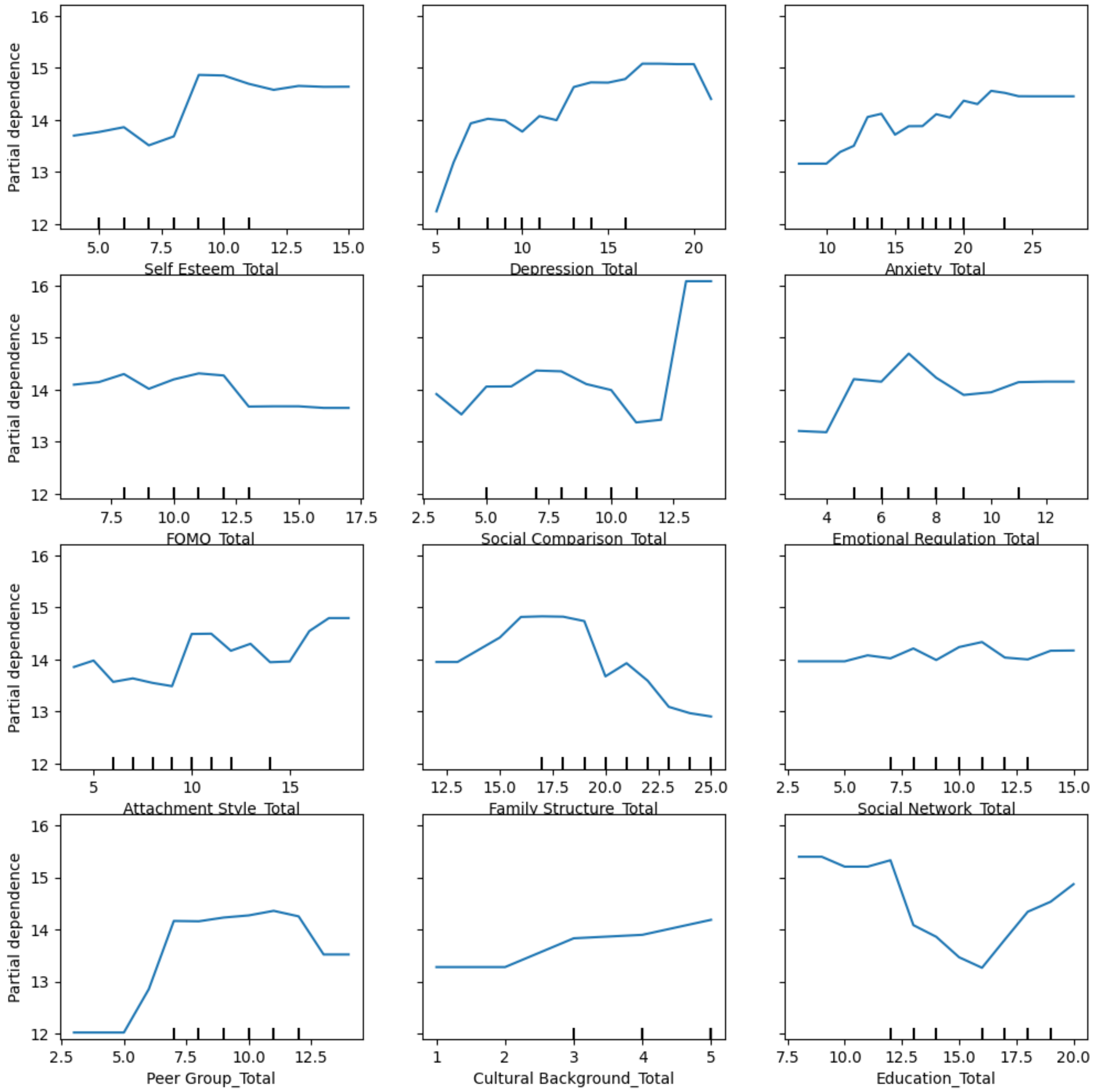


Fig. 11: Partial Dependency Plots for XGBoost

The **Partial Dependence Plots (PDPs**) given in Figure 10 and 11, illustrate the marginal effects of individual predictors on the target variable. Key factors like Depression Total and Anxiety Total show strong positive relationships, indicating significant contributions. Negative trends in Family Structure Total and Education Total suggest their inverse impact on the target variable. Other predictors, such as Cultural Background and Social Network Total, exhibit minimal effects, highlighting weaker associations.

5.4.3 *Shapley Additive explanations* (SHAP) Values

5.4.3.1 SHAP Values for Random Forest

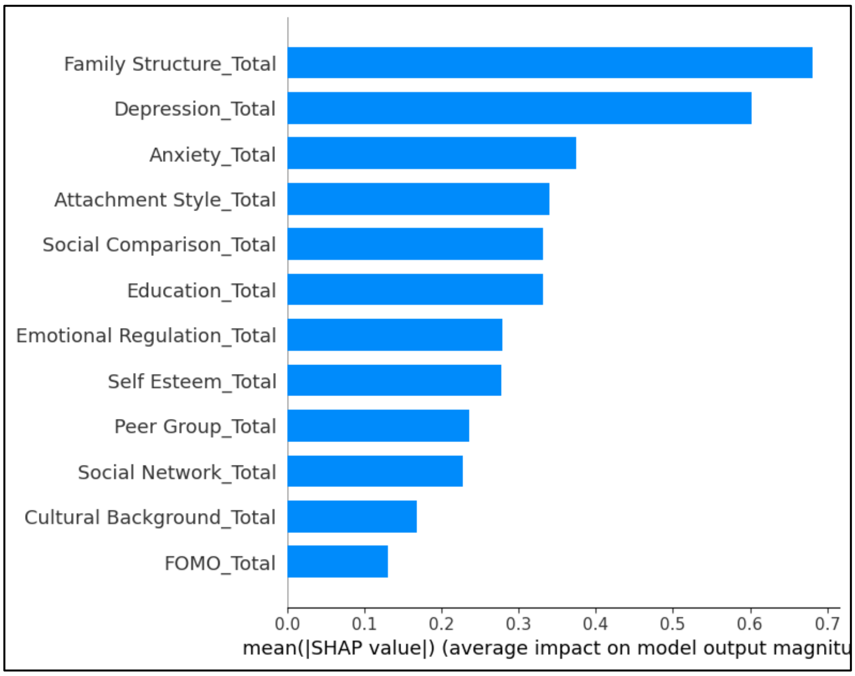


Fig. 12: SHAP Values for Decision Tree

5.4.3.2 SHAP Values for XGboost

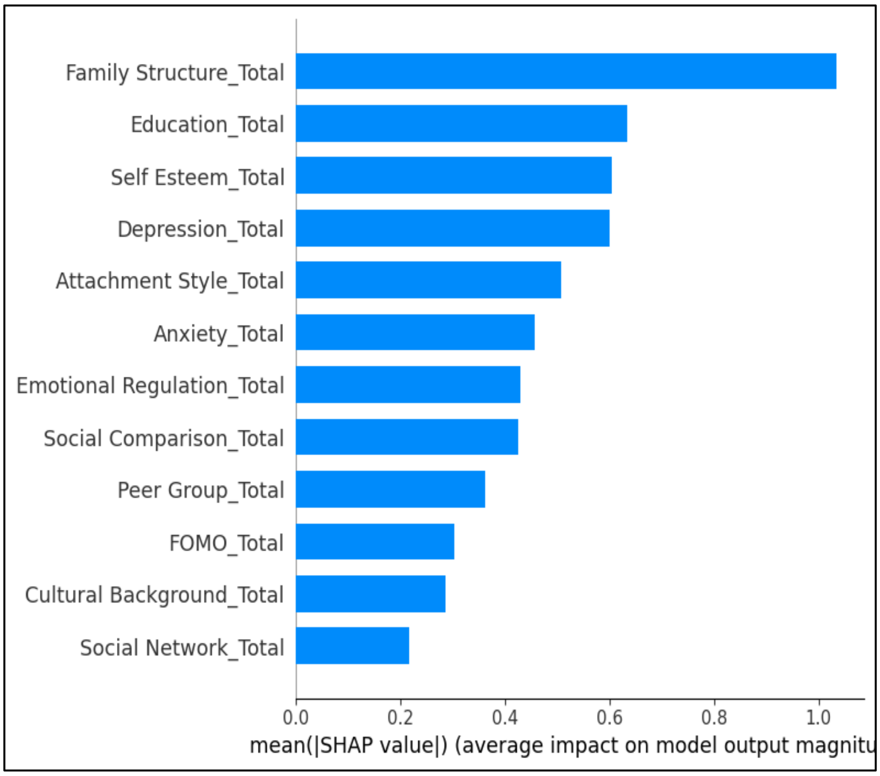


Fig. 13: SHAP Values for XGBoost

The X-axis represents the mean absolute SHAP value for each feature, which indicates the average magnitude of that feature's impact on the model's output. Features with higher values are more influential.

Across both models, Family Structure_Total consistently emerges as the most influential feature, followed Depression_Total, Anxiety_Total, Attachment style, Self Esteem, Education and Social Comparison_Total, which also have substantial impacts. Factors like Emotional Regulation_Total, Peer Group_Total, and Self Esteem_Total contribute moderately, while features such as Social Network_Total, Cultural Background_Total, and FOMO_Total have the least influence.

The figure 12 and 13, implies that mental health factors like depression, family dynamics, and anxiety play a crucial role in determining the model's predictions.

Social and emotional components such as Cultural Background influence (Cultural Background_Total), fear of missing out (FOMO_Total), and Social Network (Social Network_Total) also contribute but with less influence

Sr. No.	Model	Train RMSE	Test RMSE
1	Random Forest	1.7047	4.4016
2	XGBoost	0.0011	5.0599

Table 11: Model's Train Test RMSE Comparison

Both models show high performance on the training data, but their test RMSE values are significantly higher, indicating challenges with generalization.

Random Forest balances training and test performance better, although it may still over fit. XGBoost, despite its near-perfect training performance, does not provide a significant advantage over Random Forest on the test set, and its minimal training error suggests potential overfitting.

Further tuning of hyper parameters or techniques like cross-validation may help improve the models' test performance. We use 10-Fold-Cross Validation for further analysis.

Model	Best Parameters	Best Cross-Validation Score (MSE)
Random Forest	- max_depth = None - min_samples_split = 10 - n_estimators = 100	20.0739
XGBoost	- learning_rate = 0.01 - max_depth = 3 - n_estimators = 200 - subsample = 0.6	19.5860

Table 12: Model Comparison: Random Forest vs XGBoost

Random Forest is the marginally better performer here based on MSE, but the difference is minimal. If interpretability, training time, or deployment considerations are factors, the choice between the two may shift depending on the context. Both are strong candidates for the task.

Connection to BSMAS:

1. **Mental Health Variables (Depression and Anxiety):** These psychological factors exhibit strong positive associations with BSMAS, indicating that students experiencing emotional distress are more likely to develop social media addiction. This ties to BSMAS dimensions like mood modification and tolerance, where social media use serves as a coping mechanism for psychological challenges.

2. **FOMO and Social Comparison:** FOMO is typically associated with salience in the BSMAS framework, where individuals may prioritize social media use due to feelings of exclusion. However, the negative coefficient for FOMO suggests that alternative motivations, such as escapism, may also drive addiction. Social comparison supports the conflict dimension, as it leads to dissatisfaction and negative self-perceptions, feeling compulsive social media engagement.

3. **Attachment Styles and Family Structure:** Secure attachment styles and strong family structures are linked to lower BSMAS scores, indicating that positive interpersonal relationships reduce the dependency on social media for emotional support. This mitigates the risks of withdrawal and relapse as per the BSMAS model.

4. **Social and Peer Networks:** Offline social connections lower social media addiction by offering real-world interaction, reducing the centrality of social media in students' lives. This addresses the conflict and salience dimensions of BSMAS.

6. Conclusions

- **Key Predictors of Social Media Addiction (BSMAS):** Logistic regression identifies Self Esteem, Depression and FOMO as significant predictors, with strong positive associations indicating that emotional distress increases the likelihood of social media addiction. These findings align with BSMAS dimensions like mood modification and withdrawal, where social media use helps manage psychological challenges.
- **Role of Family and Social Structures:** Family Structure and Social Network variables show protective effects, reducing the likelihood of high BSMAS scores. Supportive family dynamics and strong offline connections decrease social media reliance, underscoring the importance of real-world relationships in reducing addiction risk.
- **Impact of Attachment Styles and Peer Influence:** Secure attachment styles and supportive peer groups correlate with lower BSMAS scores, while insecure attachment styles increase addiction risks. This reflects lower salience and conflict in users' social media engagement.
- **Social Comparison and FOMO:** Social comparison significantly contributes to social media addiction by fostering dissatisfaction and compulsive usage. FOMO shows an unexpected negative association, suggesting alternate motivations for excessive social media use in this dataset.
- **Predictive Modelling Insights:** Advanced machine learning models like Random Forest and XGBoost outperform traditional regression methods, showing the ability to capture non-linear patterns in social media addiction. However, overfitting challenges in these models indicate a need for further tuning and cross-validation.

Overall Behavioural Insights: The study reveals that social media addiction is driven by a complex mix of psychological and social factors. Mental health challenges like depression and anxiety amplify addiction risks, while strong offline relationships and healthy attachment styles act as protective factors.

7. Limitations

- **Sampling Method**: The use of random sampling might limit the generalizability of the findings, as the sample may not represent the broader population of Kolhapur city.
- **Geographical Scope**: Since data collection is restricted to Kolhapur city, the findings may not capture variations in social media addiction across different geographical contexts.
- **Dataset Size:** The dataset contains 258 observations, which may not be large enough to fully capture the complexity of the relationships among variables. A larger sample size could improve the robustness and reliability of the findings
- **Measurement Tools**: The tools or scales used to measure psychological and social factors, as well as social media addiction, may have limitations in their validity or cultural adaptability.
- **Temporal Factors**: Social media usage patterns can change over time, and the data collected may not reflect long-term trends.
- **Interview Limitations**: The interview process may introduce biases, such as social desirability bias, where participants might underreport or exaggerate their behaviors.
- **Overfitting in Machine Learning Models:** While Random Forest and XGBoost models showed exceptional performance on the training data, their high test RMSE values indicate challenges with generalization, suggesting potential overfitting. This limits the practical applicability of the models to new datasets.
- **External Validity:** The findings are specific to the given dataset and may not generalize to broader populations with different demographics, cultural backgrounds, or social media usage patterns.
- **Limited Model Comparisons:** While the analysis compares traditional statistical models and machine learning methods, it excludes other relevant algorithms such as Support Vector Machines or Neural Networks, which might provide additional perspectives.
- **BSMAS Score and Dimension Specificity:** While the Bergen Social Media Addiction Scale (BSMAS) provides an effective measure of social media addiction, it aggregates six dimensions (e.g., salience, mood modification, tolerance). The analysis does not explicitly evaluate how each dimension correlates with predictors, potentially overlooking nuanced insights.

8. Future Work

- **Alternative Modeling Approaches:** Future studies could consider using the six individual question scores from the BSMAS as predicted variables instead of the total score. This would allow the application of aggregative models to explore nuanced patterns of social media addiction.
- **Enhanced Feature Engineering:** Incorporating advanced feature selection techniques could help identify additional significant predictors of social media addiction, enhancing the robustness of machine learning models.
- **Multiclass Classification:** Instead of dichotomizing the BSMAS scores, future research could explore multiclass classification models to capture varying levels of addiction severity (e.g., mild, moderate, severe).
- **Comparison of Models:** Comparing additional machine learning algorithms, such as *Support Vector Machines* (SVM) or Neural Networks, could provide deeper insights into predictive performance and model interpretability.
- **Exploring Interaction Effects:** Investigating interaction effects between psychological and social factors in predictive modeling could reveal complex relationships not captured by current methods.

9. References

- Arora, S., Kumar, M., & Piplani, K. (2022). Social Media Addiction- Risk of Addiction in India Measured Through Bergen Social Media Addiction Scale (BSMAS). *Deleted Journal*, *22*(2), 101–113. https://doi.org/10.57198/2583-4932.1308
- Çiftci, N., & Yıldız, M. (2023). The Relationship Between Social Media Addiction, Happiness, and Life Satisfaction in Adults: Analysis with Machine Learning Approach. *International Journal of Mental Health and Addiction*, 21(5), 3500–3516. https://doi.org/10.1007/s11469-023-01118-7
- Zarate, D., Hobson, B. A., March, E., Griffiths, M. D., & Stavropoulos, V. (2022). Psychometric properties of the Bergen Social Media Addiction Scale: An analysis using item response theory. *Addictive Behaviors Reports*, *17*, 100473. https://doi.org/10.1016/j.abrep.2022.100473
- Simsek, A., Elciyar, K., & Kizilhan, T. (2019). A Comparative Study on Social Media Addiction of High School and University Students. *Contemporary Educational Technology*, *10*(2), 106–119. https://doi.org/10.30935/cet.554452