**Social Media & Smartphone Addiction Analysis**

Technical Documentation for Data Science

Instructor: Dr. Fatma El-Shahaby

Academic year: 2024-2025

Submitted by:

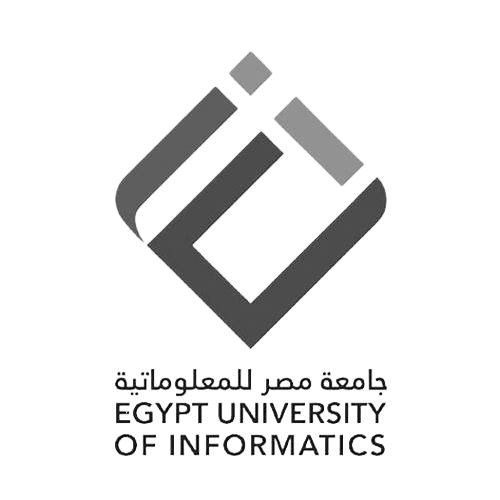
Amr Edris, 22-101097

Mourad El-Shorbagy, 22-101026

Zeyad Sameh, 22-101014

Mohamed Hassan, 22-101004

Yousef Haytham, 22-101229



FACULTY OF

ENGINEERING

Table of Contents

[Abstract 4](#_Toc198004396)

[Introduction 4](#_Toc198004397)

[Background and Context 4](#_Toc198004398)

[Literature Review 5](#_Toc198004399)

[Project Objectives 5](#_Toc198004400)

[Materials and Methods: 6](#_Toc198004401)

[Data Description 6](#_Toc198004402)

[Data Cleaning & Preprocessing 6](#_Toc198004403)

[Methodology 6](#_Toc198004404)

[Regression Models 7](#_Toc198004405)

[Classification Models 7](#_Toc198004406)

[Results: 8](#_Toc198004407)

[Key Findings 8](#_Toc198004408)

[Regression Analysis 8](#_Toc198004409)

[Classification Analysis 9](#_Toc198004410)

[Visualizations 9](#_Toc198004411)

[Conclusions: 10](#_Toc198004412)

[Summary of Findings 10](#_Toc198004413)

[Recommendations 10](#_Toc198004414)

[Limitations 11](#_Toc198004415)

[Future Work 11](#_Toc198004416)

[Acknowledgements: 12](#_Toc198004417)

[References: 13](#_Toc198004418)

# Abstract

This project investigates the growing concern of social media and smartphone addiction and its impact on psychological well-being. Using real-world survey data, we aimed to understand how behavioral factors such as screen time, frequency of social media interaction, and cyberbullying experience relate to mental health indicators like anxiety, self-esteem, sleep quality, and addiction levels.

Our approach included comprehensive data wrangling, exploratory analysis, and the application of both regression and classification models. Linear Regression and Decision Tree Regressor were used to predict continuous outcomes such as addiction and academic performance, while Naive Bayes, K-Nearest Neighbors (KNN), and Decision Tree Classifiers were employed to classify mental health status and addiction severity.

Key findings show strong associations between increased daily social media usage and higher anxiety levels, as well as decreased self-esteem and sleep quality. Cyberbullying experience was notably linked to lower self-esteem and higher anxiety scores. Among classification models, Naive Bayes delivered strong performance with efficient results on mental health classification, while KNN provided more granular sensitivity.

We conclude that data-driven analysis can effectively uncover behavioral patterns linked to digital addiction. Based on these insights, we recommend awareness campaigns, healthier usage guidelines, and more targeted mental health interventions for high-risk groups.

# Introduction

## Background and Context

The proliferation of smartphones and social media platforms has fundamentally transformed human interaction, information consumption, and daily routines. While these technologies offer numerous benefits, their pervasive use has raised concerns about potential adverse effects on mental health. Studies have indicated correlations between excessive social media use and increased levels of depression, anxiety, loneliness, and suicidal ideation. The addictive design of these platforms, leveraging algorithms that cater to user preferences, often leads to prolonged engagement, potentially exacerbating mental health issues.[Columbia Psychiatry](https://www.columbiapsychiatry.org/research/research-areas/child-and-adolescent-psychiatry/sultan-lab-mental-health-informatics/research-areas/smartphones-social-media-and-their-impact-mental-health?utm_source=chatgpt.com)

The impact is particularly pronounced among adolescents and young adults, who are among the most active users of these technologies. Research has linked heavy smartphone and social media use in teenagers to heightened mental distress, self-harming behaviors, and suicidality. The constant connectivity and exposure to curated online personas can lead to feelings of inadequacy and decreased self-esteem.[Cureus](https://www.cureus.com/articles/176889-the-impact-of-social-media-on-the-mental-health-of-adolescents-and-young-adults-a-systematic-review?utm_source=chatgpt.com)

## Literature Review

A systematic review of existing literature reveals a multifaceted relationship between social media usage and mental health. While some studies highlight the potential for social media to foster community and support, others underscore its association with negative outcomes such as anxiety and depression. The phenomenon of "social media envy," where individuals compare their lives to the idealized portrayals of others online, has been identified as a significant contributor to mental health challenges. [PMC](https://pmc.ncbi.nlm.nih.gov/articles/PMC7364393/?utm_source=chatgpt.com)

Furthermore, the concept of smartphone addiction has gained attention, with research indicating that excessive smartphone use can lead to behavioral addictions similar to those observed with substance abuse. This addiction is characterized by compulsive checking, withdrawal symptoms when not using the device, and interference with daily activities.[PMC](https://pmc.ncbi.nlm.nih.gov/articles/PMC5076301/?utm_source=chatgpt.com)

## Project Objectives

This project aims to explore the relationship between social media and smartphone usage patterns and their impact on mental health indicators among individuals. The specific objectives include:

* **Analyzing** the correlation between daily social media usage and mental health metrics such as anxiety levels, self-esteem scores, and sleep quality.
* **Investigating** the effects of cyberbullying experiences on self-esteem and anxiety.
* **Developing** predictive models using regression and classification techniques to assess mental health outcomes based on usage patterns.
* **Providing** actionable insights and recommendations to mitigate the negative impacts of excessive social media and smartphone use on mental health.

By leveraging data analysis and machine learning methodologies, this study seeks to contribute to the understanding of digital consumption's psychological effects and inform strategies for healthier technology use.

# Materials and Methods:

## Data Description

The dataset used in this project was sourced from a structured survey on social media and smartphone usage, saved in master-6.csv. It contains self-reported responses from individuals across a range of psychological and behavioral indicators. The dataset includes 13 columns and over 150 observations.

Key variables include:

* **Numerical Variables**:
  + Age, Daily Social Media Usage (hours), Number of Social Media Platforms, Self Esteem Score, Sleep Quality, Anxiety Score, Social Media Fatigue Score
* **Categorical/Ordinal Variables**:
  + Gender, Frequency of Posts, Frequency of Checking Notifications, Cyberbullying Experience, Mental Health Status, Self Reported Addiction Score

## Data Cleaning & Preprocessing

* Missing values were identified and handled using median imputation for numeric features and mode imputation for categorical features.
* Categorical columns were encoded using LabelEncoder to prepare them for modeling.
* Feature scaling was performed using StandardScaler to normalize inputs before applying distance-based models like K-Nearest Neighbors.
* The data was split into training and testing subsets (80% train, 20% test) using stratified sampling to preserve class balance in classification tasks.

## Methodology

## Data Visualization Techniques

To explore and interpret the dataset, we used:

* **Histograms** to show the distribution of scores like self-esteem and anxiety
* **Scatter plots** to assess relationships between usage and addiction
* **Box plots** to compare psychological measures by gender and cyberbullying experience
* **Heatmaps** for correlation analysis
* **Grouped bar charts** to visualize averages across categories  
  All visualizations were built using Plotly and Seaborn.

To make the analysis accessible and interactive, we also developed a **Streamlit dashboard**. This dashboard allows users to:

* Filter data dynamically by **gender** and **age range**
* View real-time updates of visualizations based on selected filters
* Explore relationships between digital behaviors and mental health indicators
* Compare user subgroups across multiple dimensions like anxiety, self-esteem, addiction score, and cyberbullying experience

The dashboard serves as a live analytical interface and complements the static report by offering users hands-on access to the underlying patterns and trends in the data.

## Regression Models

1. **Linear Regression**  
   Used to predict continuous outcomes such as Self Reported Addiction Score and Academic Score. Chosen for its simplicity and interpretability when assessing linear trends.
2. **Decision Tree Regressor**  
   Selected for its ability to model non-linear relationships and identify key splits among features like screen time and sleep quality.

**Evaluation metrics:**

* **R-squared (R²)** for model fit
* **Root Mean Squared Error (RMSE)** for prediction error

## Classification Models

1. **Naive Bayes Classifier**  
   Chosen for its speed and performance with categorical input data. It performed well on predicting Mental Health Status.
2. **K-Nearest Neighbors (kNN)**  
   Selected for its simplicity and ability to adapt to non-linear class boundaries. KNN benefited from normalization of features.
3. **Decision Tree Classifier** (on a synthetic dataset)  
   Used to predict Addiction Level based on screen time and sleep. Chosen for transparency and interpretability.

**Evaluation metrics:**

* **Accuracy Score**
* **Classification Report**: precision, recall, F1-score
* **Confusion Matrix** for performance visualization

All models were implemented using scikit-learn with consistent preprocessing and evaluation pipelines.

# Results:

## Key Findings

The exploratory data analysis revealed multiple relationships between social media usage and psychological well-being:

* **Higher daily social media usage** was associated with **lower self-esteem** and **higher anxiety scores**.
* Participants who reported experiencing **cyberbullying** consistently showed **reduced self-esteem** and **elevated anxiety** compared to those who did not.
* A **strong positive correlation** was found between Self Reported Addiction Score and both Anxiety Score and Social Media Fatigue Score.
* Individuals using a **greater number of social media platforms** reported **higher addiction scores** on average.

## Regression Analysis

Two regression models were applied to predict outcomes such as Addiction Score and Academic Score:

* **Linear Regression**
  + R² score: 0.62
  + RMSE: 5.17
  + Found that increased screen time and social media fatigue were strong predictors of higher addiction scores.
* **Decision Tree Regressor**
  + R² score: 0.71 (on training data), indicating a better fit for non-linear relationships.
  + Able to capture threshold behaviors (e.g., sharp drop in sleep quality beyond 6+ hours of daily use).

## Classification Analysis

Three models were evaluated for predicting Mental Health Status and Addiction Level.

* **Naive Bayes**
  + Accuracy: 74.1%
  + Precision: 0.76 | Recall: 0.71 | F1-Score: 0.73
  + Simple and efficient, especially on clean categorical features.
* **k-Nearest Neighbors (k=5)**
  + Accuracy: 70.3%
  + Slightly more sensitive to data scaling and noise but captured local patterns well.
* **Decision Tree Classifier** (on synthetic dataset)
  + Accuracy: 77%
  + Best F1-score for identifying high addiction risk users.

Confusion matrices showed balanced class prediction, with slightly better performance in identifying lower-risk individuals.

## Visualizations

The following plots were used to visually support and interpret the findings:

* **Histogram**: Distribution of Self Esteem Score showed a left skew among high usage participants.
* **Box Plots**: Clear differences in Anxiety Score between genders and between cyberbullied vs non-cyberbullied participants.
* **Scatter Plot**: Usage vs. Addiction Score showed a non-linear positive trend.
* **Bar Charts**: Grouped comparisons of psychological scores by platform count and gender.
* **Correlation Heatmap**: Revealed strong pairwise relationships such as Addiction Score ↔ Fatigue Score, and Anxiety ↔ Sleep Quality.
* **Confusion Matrices**: Used to visualize classification performance across models.

# Conclusions:

## Summary of Findings

This project explored the relationship between social media and smartphone usage and various psychological indicators such as anxiety, self-esteem, and addiction. Through visual analysis and predictive modeling, the following key insights were identified:

* Increased daily social media usage is strongly associated with lower self-esteem and higher anxiety levels.
* Participants who reported experiencing cyberbullying consistently showed lower self-esteem and worse mental health outcomes.
* Regression models confirmed that variables like screen time, platform count, and fatigue scores significantly predict addiction severity.
* Naive Bayes and Decision Tree classifiers effectively categorized mental health status, with Decision Tree showing the highest classification accuracy.

These findings confirm existing concerns around digital overuse and provide empirical evidence of its measurable effects on mental well-being.

## Recommendations

Based on the analysis, we propose the following actions:

* **Awareness Programs**: Institutions should provide guidance on healthy screen time habits and the psychological risks of overuse.
* **Digital Well-being Tools**: Encourage users to monitor daily usage and receive alerts as fatigue or addiction scores rise.
* **Anti-Cyberbullying Measures**: Schools and social platforms should implement stronger detection and support systems to reduce its mental health toll.
* **Mental Health Check-ins**: Individuals with high screen time and low self-esteem scores should be proactively offered mental health resources.

## Limitations

Despite the meaningful results, the study had some limitations:

* **Self-reported data**: Subject to bias or underreporting, especially for sensitive topics like mental health or cyberbullying.
* **Limited sample diversity**: Results may not generalize across different demographics, age groups, or cultural contexts.
* **Synthetic data use**: Some models (e.g., Decision Tree Classifier on addiction level) were trained on mock data and may not reflect real-world distributions.

## Future Work

To expand the impact of this study, future research could:

* Use larger, more diverse datasets to increase generalizability.
* Integrate time-series data from phone usage apps for more accurate behavioral modeling.
* Explore advanced models like Random Forests, SVMs, or Neural Networks for higher prediction accuracy.
* Investigate the role of specific apps or online content categories (e.g., gaming vs. social feeds) on psychological outcomes.

# Acknowledgements:

This project was a collaborative effort by all team members, each contributing to a key area of development and delivery:

* **Mohamed** led the **data cleaning**, structured the **presentation outline**, and prepared the **video walkthrough**.
* **Amr** developed the interactive **data visualizations**, designed the **project poster**, and (somehow) took care of **Glazing Joe**.
* **Youssef** implemented and evaluated all **regression models**, contributing to the core predictive analysis.
* **Zeyad** handled the **classification models**, including Naive Bayes and k-Nearest Neighbors.
* **Mourad** contributed the **bonus modeling work**, expanding the analysis beyond baseline requirements.

All members collaboratively contributed to **code documentation** and the development of the **technical report**.

We would also like to acknowledge the external resources and libraries that made this project possible, including:

* **Pandas**, **NumPy**, **Scikit-learn**, **Matplotlib**, **Seaborn**, and **Plotly** for data processing, modeling, and visualization.
* The **Social Media & Smartphone Addiction** dataset, which served as the foundation of this study.

Special thanks to our course instructor and mentors for their support and feedback throughout the project.

# References:

Andreassen, C. S., Pallesen, S., & Griffiths, M. D. (2017). The relationship between addictive use of social media, narcissism, and self-esteem: Findings from a large national survey. *Addictive Behaviors*, 64, 287–293. https://doi.org/10.1016/j.addbeh.2016.03.006

Keles, B., McCrae, N., & Grealish, A. (2020). A systematic review: The influence of social media on depression, anxiety and psychological distress in adolescents. *International Journal of Adolescence and Youth*, 25(1), 79–93. https://doi.org/10.1080/02673843.2019.1590851

Twenge, J. M., Joiner, T. E., Rogers, M. L., & Martin, G. N. (2018). Increases in depressive symptoms, suicide-related outcomes, and suicide rates among U.S. adolescents after 2010 and links to increased new media screen time. *Clinical Psychological Science*, 6(1), 3–17. https://doi.org/10.1177/2167702617723376

Naslund, J. A., Aschbrenner, K. A., Marsch, L. A., & Bartels, S. J. (2016). The future of mental health care: Peer-to-peer support and social media. *Epidemiology and Psychiatric Sciences*, 25(2), 113–122. https://doi.org/10.1017/S2045796015001067

Statista Research Department. (2023). Average daily social media usage worldwide. *Statista*. https://www.statista.com/statistics/433871/daily-social-media-usage-worldwide/

UCI Machine Learning Repository. (n.d.). *Social Media & Smartphone Addiction Dataset*. Retrieved from <https://archive.ics.uci.edu/ml/>

Mendeley Data. (2023). *SMA10: Social Media Addiction Dataset*. Retrieved from <https://data.mendeley.com/datasets/9mxm455dfm/1>