Sentiment Analysis of Social Media Posts

Jaqueline Vallejo Hinojosa

Abstract

Social media plays a pivotal role in shaping the daily lives of teenagers and young adults, influencing their emotions and behaviors. This research paper explores the extent of this influence by analyzing the prevalence of positive, negative, and neutral content encountered by young adults on various social platforms. Special emphasis is placed on negative content to assess its specific impacts on mental health and social interactions. The findings seek to contribute to ongoing discussions about digital well-being and provide insights for policymakers and educators on managing social media's impact on young populations.

1 Introduction

In the digital age, social media has become a global force, influencing nearly every aspect of daily life, particularly for teenagers and young adults. Platforms such as Instagram, Facebook, and Twitter present a wide array of content, ranging from the positive to the negative. While social media is a crucial source of information and social interaction, its impact on mental health and emotional well-being is a persistent concern. Numerous studies have documented its negative effects on teenagers and young adults. This raises an important question: Can identifying and reducing exposure to negative content improve mental health outcomes?

The primary objective of this research is to quantify and compare the proportions of positive, negative, and neutral content that young adults encounter daily on social media, and to examine whether reducing exposure to negative content yields positive mental health benefits. Understanding these dynamics is essential for developing interventions and policies to safeguard the mental health of young adults in the digital realm.

The paper is organized as follows: It begins with a literature review, followed by the methodology employed to analyze the dataset and identify the predominant types of content. The results are then presented, and insights and implications for future policies and practices are discussed. The paper concludes by summarizing the key findings

2 Literature Review

Over the past decade, social media has profoundly influenced young adult life, offering unprecedented connectivity but often at a substantial cost to mental health. In 2024, Kathy Katella addressed this issue in her guide aimed at parents, discussing both the benefits and risks of social media for teenagers. Her work highlights the dual nature of digital interactions: while they facilitate connections with a diverse network of peers, they also expose young users to a significant amount of harmful content. Social media itself isn’t inherently detrimental to teens, but its excessive use and the presence of inappropriate, harmful content can lead to more severe problems like eating disorders, self-harm, and even suicide.

Katella particularly notes that the adolescent brain, which is highly sensitive between the ages of 10-19, is undergoing crucial developmental stages in identity formation and self-worth. The prevalent use of social media during this vulnerable period can thus have detrimental effects, depending on the nature of the content encountered. Her findings underscore the increased anxiety, disrupted self-esteem, and issues such as body image concerns and disordered eating behaviors, particularly among young girls. These problems are compounded by poor sleep quality, often directly linked to excessive social media usage.

Our research employs sentiment analysis to quantify the exposure of teenagers to negative content on social media. By identifying the scale of negative influences, we aim to inform and develop policies that enable parents to better safeguard their children from these adverse effects, thereby mitigating the psychological risks posed by social media.

3 Methodology

This section details the methodologies employed to analyze the impact of social media content on the mental health of young adults. It specifically focuses on quantifying the exposure to negative content across major social media platforms, including Facebook, Instagram, and Twitter. The approach is designed to measure how different types of content influence mental well-being and to identify potential correlations between content sentiment and psychological effects. This section also outlines the specific strategies and tools used for data collection, processing, and analysis, ensuring a comprehensive understanding of the operational framework and the analytical techniques applied in this study.

4 Data Acquisition and Processing

The dataset utilized in this study was sourced from Kaggle, a platform renowned for its diverse array of publicly available datasets provided by users and organizations. We specifically selected this dataset due to its extensive compilation of social media posts from popular platforms such as Facebook, Instagram, and Twitter, which are predominantly used by young adults.

The dataset comprises 732 posts collected over the period from January 2012 to September 2023. Each entry contains the post's text, the date and time of posting, and anonymized user engagement metrics, including likes and retweets. This breadth of data is particularly pertinent to our study’s objective of analyzing the impact of social media content on the mental health of young adults, as it offers a rich spectrum of user-generated content suitable for detailed sentiment analysis and longitudinal trend observations across various platforms.

Upon acquisition, significant efforts were made to preprocess the dataset to ensure its suitability for detailed analysis. The preprocessing steps included stripping the text of stopwords and punctuation, converting all text to lowercase, and tokenizing the text. These measures were critical for enhancing the analytical accuracy and for facilitating the detection of frequently occurring content themes through subsequent frequency analysis.

5 Text Analysis Techniques

To effectively analyze the content of social media posts and their potential impact on mental health, several text analysis methods were employed: Binary, Term Frequency (TF), and Term Frequency-Inverse Document Frequency (TF-IDF). These methods are specifically chosen to quantify the presence and significance of words within the dataset, thereby enabling a nuanced understanding of content exposure among young adults. The Binary method assesses the presence or absence of terms; Term Frequency measures the frequency of terms within individual posts, reflecting their prominence; and TF-IDF evaluates the relative importance of terms by considering their frequency across all documents, helping to highlight words that are particularly significant in specific contexts. Detailed descriptions of each method will be provided in the following sections, offering insights into their application and the specific value they bring to this study.

5.1 Binary

The binary text representation model was employed to determine the presence or absence of specific terms in each social media post, allowing for the detection of distinct types of content. This method helps to identify potentially harmful or beneficial content across the dataset by focusing solely on word occurrence rather than frequency. By representing each word as either present (1) or absent (0), this approach enables a structured analysis of how often key terms related to mental health appear in social media discussions, without being influenced by repeated occurrences within a single post.

5.2 Term Frequency

The Term Frequency (TF) technique was used to measure the importance of words within a document by calculating their frequency relative to the total number of words in that document. Unlike the binary text representation model, which only accounts for the presence or absence of words, TF quantifies word occurrences, allowing for a more detailed assessment of prominent themes and topics within individual social media posts. This technique helps identify frequently occurring terms, which can be correlated with user engagement metrics such as likes and retweets, providing insights into content trends and user interactions across different platforms. Figures 1–3 present the word clouds for positive, negative, and neutral sentiments as detected by the model.

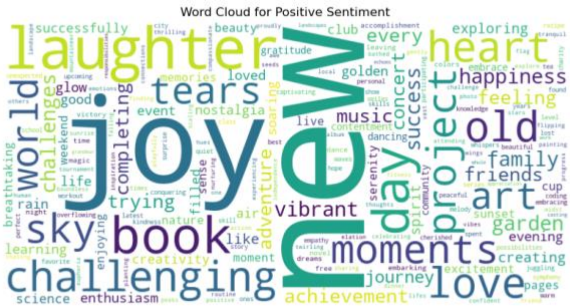

Figure 1: Word Cloud of Positive Sentiment

Figure 2: Word Cloud of Negative Sentiment

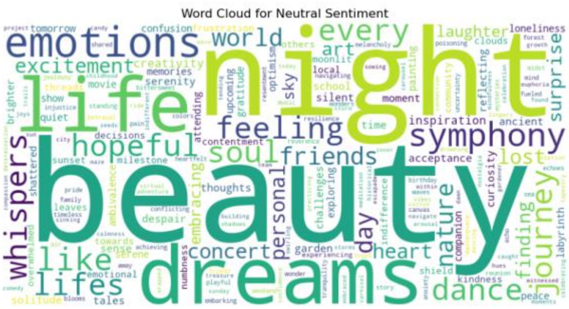

Figure 3: Word Cloud of Neutral Sentiment

5.3 Term Frequency + Inverse Document Frequency

The Term Frequency-Inverse Document Frequency (TF-IDF) technique was applied to assess the importance of words within social media posts while minimizing the influence of commonly used terms. Unlike Term Frequency (TF), which measures how often a word appears in a document, TF-IDF adjusts for words that frequently occur across multiple documents, ensuring that distinctive and meaningful terms are emphasized. The TF-IDF value increases proportionally to the number of times a word appears in a document but is offset by its frequency across the entire dataset, thereby prioritizing terms that hold greater contextual significance rather than those that appear commonly across posts.

This method was particularly effective in distinguishing high-impact words related to mental health concerns, such as terms associated with anxiety, depression, and positivity. By highlighting emotionally significant words, TF-IDF played a crucial role in sentiment analysis and content evaluation.

Employing these text analysis methods facilitated a comprehensive assessment of social media content, enabling a deeper understanding of its emotional tone and thematic significance. These insights help to illustrate the potential influence of social media content on young adults' mental health and contribute to identifying patterns in online discourse.

6 Results

This section presents the key findings obtained through sentiment analysis and text-processing techniques, including Binary text representation, Term Frequency (TF), and Term Frequency-Inverse Document Frequency (TF-IDF). These methods were employed to identify patterns in social media content and assess its potential impact on young adults' mental health.

The results are structured as follows: First, we examine the sentiment distribution of social media posts, highlighting the proportion of positive, negative, and neutral content. Next, we explore the most frequently occurring terms across different sentiment categories using Term

Frequency analysis. Finally, we analyze the TF-IDF scores to determine the most contextually significant words in the dataset, providing deeper insights into recurring themes in mental health discussions on social media.

6.1 Sentiment Distribution

To understand the overall sentiment distribution in our dataset, we calculated the proportion of posts classified as negative, positive, and neutral. The results are as follows:

- Negative Sentiment: 17.21% (0.1721)
- Positive Sentiment: 38.52% (0.3852)
- Neutral Sentiment: 44.26% (0.4426)

The results indicate that negative posts make up a significant portion of social media content, though they are outnumbered by both neutral and positive posts. However, given the potential psychological impact of negative content, further analysis was conducted to assess how negative sentiment is distributed across various factors, such as time of day, user engagement (likes/retweets), geographic distribution, and long-term trends.

To explore the impact of negative posts in our dataset, we analyzed various dimensions of sentiment distribution using the following visualizations:

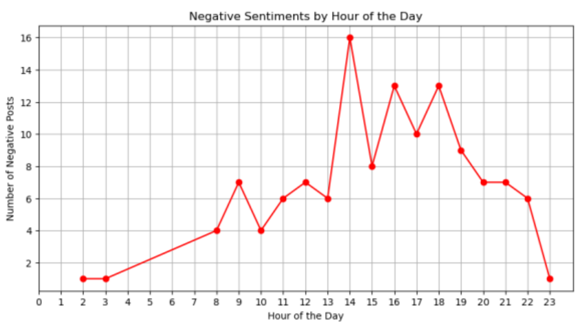

Figure 4: Negative Sentiments by Hour of Day

A time-based analysis was conducted to determine when negative posts are most frequently shared throughout the day. **Figure 4** presents the distribution of negative sentiments by hour, revealing a peak at 14:00 (2:00 PM) with 16 negative posts**.** Additional spikes are observed at 16:00 (4:00 PM) and 18:00 (6:00 PM), each recording 13 negative posts**.** This graph helps identify potential patterns in negative content activity, suggesting that negative sentiment is more prevalent during specific hours of the day.

Scatter plots illustrate the relationship between sentiment scores and user engagement metrics (likes and retweets). This analysis helps assess whether negative posts receive higher or lower interaction compared to positive and neutral content. **Figures 5 and 6** reveal a weak correlation between sentiment and user engagement, as indicated by the scattered distribution of data points forming vertical patterns. However, there are noticeably fewer data points for negative sentiments, suggesting lower engagement with negatively classified posts.

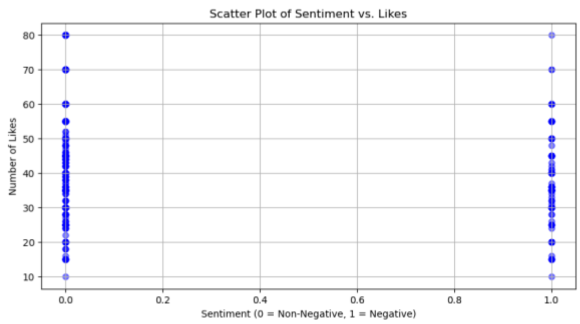

Figure 5: Scatter Plot Sentiment vs. Likes

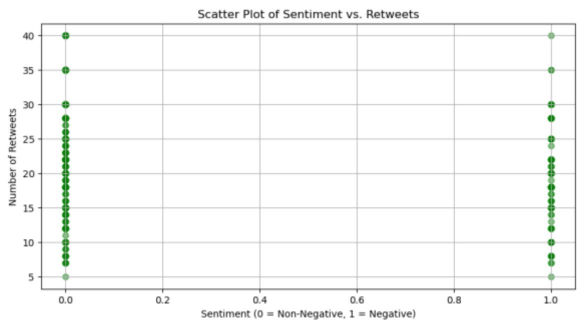

Figure 6: Scatter Plot Sentiment vs. Retweets

Bar graph compares the number of negative posts across different countries, highlighting regional variations in social media sentiment. As seen in **Figure 7,** the USA has the highest number of negative posts (above 30), followed by Canada (25), while Spain records the lowest (fewer than 3 negative posts). This geographic breakdown offers insights into how negative sentiment varies across different cultural and social contexts.

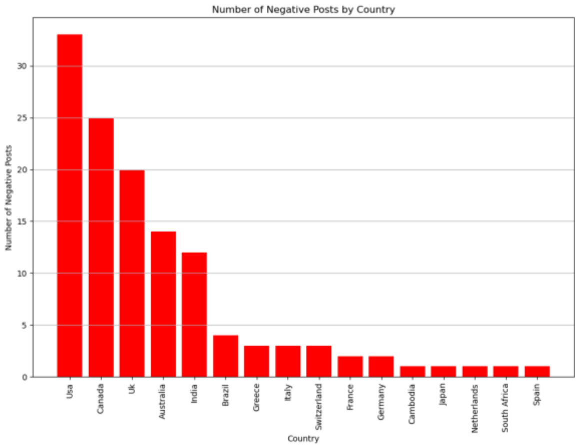

Figure 7: Negative Posts by Country

The time-series graph analyzes trends in negative posts over time, assessing whether negative sentiment has increased or decreased from 2012 to 2023. This visualization highlights fluctuations and notable spikes in online negativity throughout the years. **From** 2020 to 2022**,** the graph reveals a series of smaller patterns, culminating in a peak in 2023**,** marking the highest recorded level of negative sentiment.

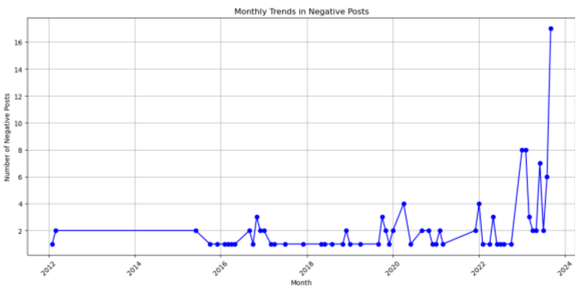

Figure 8: Monthly Trends in Negative Posts

By examining temporal, engagement-based, geographic, and historical factors, this analysis provides deeper insights into how negative sentiment manifests on social media and its potential impact on young adults' mental health. The following sections will further explore the Binary, Term Frequency (TF) results and TF-IDF scores to identify key patterns in content distribution and significance.

6.2 Binary Results

The Binary Model emerged as the best overall performer, achieving the highest test accuracy (77.0%) and AUROC (0.8473), making it the most reliable for sentiment classification. It demonstrated perfect precision (1.00) for negative sentiment, ensuring no false positives, and showed strong recall for neutral sentiment (0.91). However, recall for negative (0.60) and positive (0.68) sentiments was slightly lower, indicating that some posts in these categories were misclassified. Additionally, the model tended to over-classify neutral sentiment, leading to slightly lower precision for this category. Despite these minor issues, the Binary Model remains the most balanced and effective choice for sentiment classification in this study.

6.3 Term Frequency Results

The TF Model performed slightly below the Binary Model, with 74.3% accuracy and AUROC of 0.8365. It showed decent recall across all sentiment classes, with values of 0.60 for negative**,** 0.88 for neutral, and 0.65 for positive sentiment**.** However, its accuracy and classification balance were weaker than the Binary Model, making it a solid alternative but not the top choice**.**

6.4 TF-IDF Scores

The TF-IDF Model, in contrast, performed the weakest, with an accuracy of 64.9% and AUROC of 0.8174. The model particularly struggled with negative sentiment detection, with an extremely low recall of 0.10, meaning that 90% of actual negative posts were misclassified. While TF-IDF was effective at identifying neutral sentiment (recall = 0.91), its overall classification accuracy and balance were poor, making it unsuitable for direct sentiment classification. However, TF-IDF remains valuable for feature extraction and identifying key words in sentiment analysis rather than for classification itself.

Overall, the Binary Model is the best choice for sentiment classification, offering the highest accuracy and a good balance between precision and recall. The TF Model is a viable backup option, performing better than TF-IDF but not as well as Binary. TF-IDF, while weak for classification, may still be useful for deeper content analysis to identify important terms contributing to sentiment. Based on these results, the Binary Model is recommended as the primary method for sentiment classification in this study.

7 References

Katella, Kathy, “How Social Media Affects Your Teen’s Mental Health: A Parent’s Guide”, Yale Medicine, June 17, 2024 https://www.yalemedicine.org/news/social-media-teen-mental-health-a-parents-guide, Accessed March 16, 2025.

Kashish Parmar, “Social Media Sentiments Analysis Dataset’, Kaggle, 2024, https://www.kaggle.com/datasets/kashishparmar02/social-media-sentiments-analysis-dataset, Accessed March 16, 2025.

code GitHub: https://github.com/rsm-j1vallej/MGTA415_Final_Project.git