|  |
| --- |
| **Sentiment Analysis of Social Media Posts** |
|  |
| **Jaqueline Vallejo Hinojosa** |
|  |
|  |
|  |
|  |

Abstract

|  |
| --- |
| 000  001  002  003  004  005  006  007  008  009  010  011  012  013  014  015  016  017  018  019  020  021  022  023  024  025  026  027  028  029  030  031  032  033  034  035  036  037  038  039  040  041  042  043  044  045  046  047  048  049 |

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.

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

Literature Review

|  |
| --- |
| 050  051  052  053  054  055  056  057  058  059  060  061  062  063  064  065  066  067  068  069  070  071  072  073  074  075  076  077  078  079  080  081  082  083  084  085  086  087  088  089  090  091  092  093  094  095  096  097  098  099 |

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.

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.

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.

|  |
| --- |
| 100  101  102  103  104  105  106  107  108  109  110  111  112  113  114  115  116  117  118  119  120  121  122  123  124  125  126  127  128  129  130  131  132  133  134  135  136  137  138  139  140  141  142  143  144  145  146  147  148  149 |

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.

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.

## **Binary**

|  |
| --- |
| 150  151  152  153  154  155  156  157  158  159  160  161  162  163  164  165  166  167  168  169  170  171  172  173  174  175  176  177  178  179  180  181  182  183  184  185  186  187  188  189  190  191  192  193  194  195  196  197  198  199 |

Papers that have been or will be submitted to other meetings or publications must indicate this at submission time in the START submission form, and must be withdrawn from the other venues if accepted by ACL 2020. Authors of papers accepted for presentation at ACL 2020 must notify the program chairs by the camera-ready deadline as to whether the paper will be presented. We will not accept for publication or presentation the papers that overlap significantly in content or results with papers that will be (or have been) published elsewhere.

Authors submitting more than one paper to ACL 2020 must ensure that submissions do not overlap significantly (>25%) with each other in content or results.

## **Term Frequency**

Manuscripts must be in two-column format. Exceptions to the two-column format include the title, authors' names and complete addresses, which must be centered at the top of the first page, and any full-width figures or tables (see the guidelines in Section 7.10). **Type single-spaced.** Start all pages directly under the top margin. The manuscript should be printed single-sided and its length should not exceed the maximum page limit described in Section 4. Pages should be numbered in the version submitted for review, but **pages should not be numbered in the camera-ready version**.

A close up of words

AI-generated content may be incorrect.

Figure 1: Word Cloud of Positive Sentiment

A close up of words

AI-generated content may be incorrect.

Figure 2: Word Cloud of Negative Sentiment

A close up of words

AI-generated content may be incorrect.

Figure 3: Word Cloud of Neutral Sentiment

## **Term Frequency + Inverse Document Frequency**

For the production of the electronic manuscript you must use Adobe's Portable Document Format (PDF). Please make sure that your PDF file includes all the necessary fonts (especially tree diagrams, symbols, and fonts with Asian characters). When you print or create the PDF file, there is usually an option in your printer setup to include none, all or just non-standard fonts. Please make sure that you select the option of including ALL the fonts. **Before sending it, test your PDF by printing it from a computer different from the one where it was created.** Moreover, some word processors may generate very large PDF files,

Results

|  |
| --- |
| 200  201  202  203  204  205  206  207  208  209  210  211  212  213  214  215  216  217  218  219  220  221  222  223  224  225  226  227  228  229  230  231  232  233  234  235  236  237  238  239  240  241  242  243  244  245  246  247  248  249 |

For reasons of uniformity, Adobe's **Times Roman** font should be used. If Times Roman is not available, you may use the Times New Roman or Computer Modern Roman.

Table 1 specifies what font sizes and styles must be used for each type of text in the manuscript.

A printed ruler (line numbers in the left and right margins of the article) should be presented in the version submitted for review, so that reviewers may comment on particular lines in the paper.

A graph with red lines

AI-generated content may be incorrect.

Figure 4: Negative Sentiments by Hour of Day

A graph with lines in it

AI-generated content may be incorrect.

Figure 5: Scatter Plot Sentiment vs. Likes

A graph with lines in it

AI-generated content may be incorrect.

Figure 6: Scatter Plot Sentiment vs. Retweets

A graph of the number of negative posts by country

AI-generated content may be incorrect.

Figure 7: Negative Posts by Country

A graph with blue lines

AI-generated content may be incorrect.

Figure 8: Monthly Trends in Negative Posts

References

|  |
| --- |
| 250  251  252  253  254  255  256  257  258  259  260  261  262  263  264  265  266  267  268  269  270  271  272  273  274  275  276  277  278  279  280  281  282  283  284  285  286  287  288  289  290  291  292  293  294  295  296  297  298  299 |

|  |
| --- |
| 300  301  302  303  304  305  306  307  308  309  310  311  312  313  314  315  316  317  318  319  320  321  322  323  324  325  326  327  328  329  330  331  332  333  334  335  336  337  338  339  340  341  342  343  344  345  346  347  348  349 |

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.

|  |
| --- |
| 350  351  352  353  354  355  356  357  358  359  360  361  362  363  364  365  366  367  368  369  370  371  372  373  374  375  376  377  378  379  380  381  382  383  384  385  386  387  388  389  390  391  392  393  394  395  396  397  398  399 |

|  |
| --- |
| 400  401  402  403  404  405  406  407  408  409  410  411  412  413  414  415  416  417  418  419  420  421  422  423  424  425  426  427  428  429  430  431  432  433  434  435  436  437  438  439  440  441  442  443  444  445  446  447  448  449 |

|  |
| --- |
|  |
|  |

|  |
| --- |
| 450  451  452  453  454  455  456  457  458  459  460  461  462  463  464  465  466  467  468  469  470  471  472  473  474  475  476  477  478  479  480  481  482  483  484  485  486  487  488  489  490  491  492  493  494  495  496  497  498  499 |

|  |
| --- |
| 500  501  502  503  504  505  506  507  508  509  510  511  512  513  514  515  516  517  518  519  520  521  522  523  524  525  526  527  528  529  530  531  532  533  534  535  536  537  538  539  540  541  542  543  544  545  546  547  548  549 |