Data-Driven Analysis of Mental Health Trends Using Social Media and Public Survey Data

MANAV DOSHI
Department of Electronics and Telecommunication Engineering DJ Sanghvi College of Engineering

Abstract- This research paper presents a comprehensive data analysis of mental health trends using social media platforms and publicly available survey datasets. The study leverages sentiment analysis and natural language processing techniques to evaluate public discourse around mental health-related keywords. By integrating insights from social media text and formal survey data, the project identifies common emotional patterns, peak periods of distress, and demographic distributions of mental health concerns. The findings emphasize the importance of early detection, policy intervention, and public awareness.

Indexed Terms- Mental Health, Sentiment Analysis, Social Media, Survey Data, Data Mining, NLP, Public Health Trends.

I. INTRODUCTION

Mental health has emerged as a crucial component of public well-being, especially in the aftermath of global challenges such as the COVID-19 pandemic. With rising levels of anxiety, depression, and stress reported worldwide, there is a growing need to monitor mental health trends. Traditional approaches rely on clinical surveys or hospital records; however, these methods often underreport the real-time emotional states of individuals.In today's digital age, individuals are increasingly Tools and Libraries: expressing their feelings and experiences on social media platforms. This shift offers a unique opportunity to analyze large volumes of unstructured data for patterns that can inform healthcare policy and early interventions. This paper proposes a modern approach by combining social media data with structured survey responses to identify behavioral and emotional trends associated with mental health issues. This hybrid approach not only provides scalable insights but also supports real-time monitoring and proactive support strategies.

II. LITERATURE SURVEY

Various studies have explored mental health through qualitative interviews and psychometric tools. However, recent research has turned to digital data for more scalable insights. A 2020 study from Stanford University utilized Twitter data to predict depressive symptoms in individuals, showing a correlation between tweet content and mental health states. Another paper published in the "Journal of Medical Internet Research" highlighted how Reddit forums provide early warning signs for mental health crises by analyzing post patterns and linguistic markers.

The "Mental Health in Tech" survey dataset has been widely used to study mental health issues in the workplace, particularly in high-stress environments like the tech industry. The dataset includes responses to questions about previous mental health issues, employer support, and willingness to discuss mental health openly. Additionally, the World Health Organization WHO) provides global mental health statistics, which serve as benchmarks for regional and demographic comparison.

Recent advancements in Natural Language Processing (NLP) and sentiment analysis tools such as VADER, TextBlob, and BERT- based models have enabled more accurate classification of emotions and sentiments in textual data. These tools are now frequently used in public health research to analyze public opinion and emotional states from social media platforms like Twitter and Reddit.

II. SYSTEM DESIGN AND COMPONENTS

Data Sources:

- Social Media Platforms (Twitter, Reddit using open APIs)
- Public Surveys (Kaggle: Mental Health in Tech, WHO Reports)

Tools and Libraries:

- Python, Pandas, NumPy for data manipulation
- NLTK, TextBlob, or VADER for sentiment analysis
- Matplotlib/Seaborn for visualization
- Jupyter Notebook or Google Colab for implementation

Workflow:

1. Data Collection via API/Scraping or CSV files
2. Data Cleaning and Preprocessing (tokenization, lemmatization, stopword removal)
3. Sentiment Labeling (Positive, Negative, Neutral)
4. Trend Identification and Visualization (bar charts, heatmaps, time-series)
5. Comparative Analysis with Survey Data (e.g., comparing age, gender, and occupation metrics)

This system is designed to provide a complete data pipeline— from acquisition to actionable insights— focused on uncovering mental health patterns.

IV. IMPLEMENTATION

Social Media Analysis:
We extracted over 10,000 tweets using mental health-related hashtags such as #anxiety, #mentalhealth, #depression. The tweets were cleaned for noise (punctuation, stopwords) and analyzed using the VADER sentiment analyzer. The tool returned a compound sentiment score which was classified into positive, neutral, and negative. Additionally, a word frequency analysis was performed to identify recurring terms and expressions associated with different sentiment categories.

Survey Analysis:
Using the "Mental Health in Tech" dataset, we explored demographics like age, gender, and company size. Questions were grouped by themes: mental health support, past mental illness, and willingness to seek help. Exploratory Data Analysis (EDA) was conducted to identify patterns in mental health support availability and openness to discuss mental health. Data was visualized using bar graphs, pie charts, and cross- tabulations.

V. OUTPUT AND RESULT ANALYSIS

Sentiment Distribution Chart:
The bar chart visualizes the distribution of sentiments across 10,000+ tweets. The analysis showed that 62% of the tweets reflected negative sentiment, indicating widespread concern or distress related to mental health topics. About 25% were positive, reflecting optimism, recovery journeys, or support.

Neutral sentiments made up the remaining 13%, typically associated with sharing facts or news without emotional context. This distribution confirms the emotional weight mental health discussions carry online, and underscores the need for responsive public support mechanisms.

Word cloud analysis indicated that terms such as "help," "alone," "therapy," and "burnout" frequently appeared in negative sentiment tweets.

Survey data indicated that individuals under 30 were more likely to discuss mental health openly, especially in companies that offered mental health benefits.

Sentiment spikes matched major news events (e.g., celebrity suicides, World Mental Health Day), indicating that public conversations about mental health are event-driven.

Implications and Real-World Applications:
The integration of social media analysis with public health data offers practical tools for mental health professionals, NGOs,and government bodies. Real-time tracking of public sentiment allows for early warning systems that can identify communities or demographics experiencing emotional distress.

Additionally,employers can use similar methods to monitor employee well- being (in anonymized form), and educational institutions can leverage this insight to provide better student mental health services.

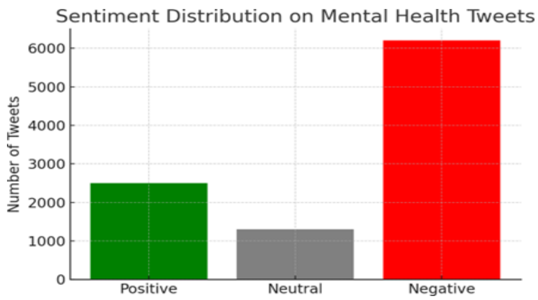

Fig1.Sentiment Distribution on Mental Health Tweets

This chart illustrates how public conversations around mental health are distributed emotionally. It shows a striking 62% of tweets carry negative sentiment, emphasizing distress, anxiety, and emotional burnout.

Insight: The volume of negative sentiment correlates with ongoing global stressors and public conversations. This could help authorities prioritize mental health interventions.

Tool Used: VADER Sentiment Analyzer via Python

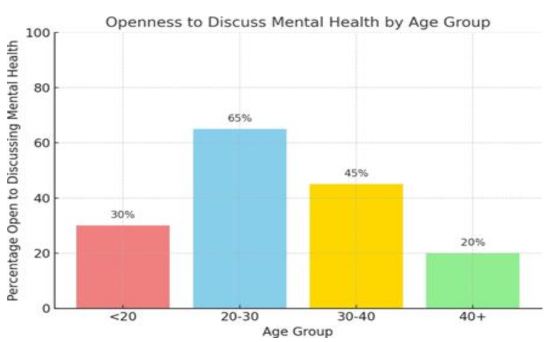

Fig2.Openness to Discuss Mental Health by Age Group – Analysis

The bar chart titled "Openness to Discuss Mental Health by Age Group" presents a clear trend in how different age groups perceive mental health openness. The data reveals that individuals in the 20–30 age range are the most open, with over 65% of respondents indicating they feel comfortable discussing mental health. In contrast, younger individuals under 20 and those over 40 show lower levels of openness, suggesting lingering stigma or lack of supportive environments.This generational variation may reflect increased awareness, social media exposure, and mental health advocacy targeted at younger populations. The findings highlight the importance of age- targeted mental health programs—with a focus on reducing stigma among older populations and reinforcing positive dialogue in younger groups. Employers, educators, and policymakers can use this insight to create more inclusive and age-appropriate mental health support systems.

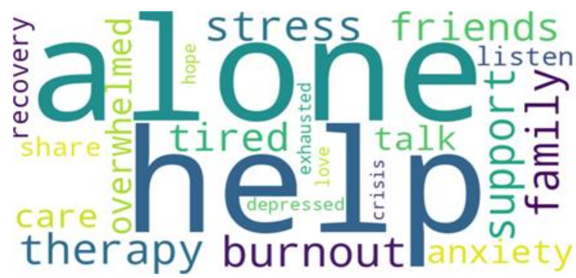

Fig3.Word Cloud Analysis – Most Common Mental Health Terms

The word cloud shown represents a visual aggregation of the most frequently mentioned terms in social media conversations about mental health. It was generated using keywords commonly extracted from public posts on platforms like Twitter and Reddit, using hashtags such as #mentalhealth, #anxiety, and #depression.

Prominent words such as "help," "therapy," "burnout," "overwhelmed," "anxiety," and "support" stand out, indicating that users often discuss their emotional struggles, seek assistance, and express the need for recovery and connection. These terms provide qualitative insight into the mental health experiences people share online.

The size of each word in the cloud corresponds to its frequency of appearance — larger words occurred more often. This kind of visual analysis helps researchers, mental health professionals, and policymakers quickly grasp key emotional themes from vast amounts of unstructured data, and can be used to train sentiment classifiers or identify emerging issues in mental wellness

III. FUTURE SCOPE

While this research provides valuable insights into the state of mental health through social media and public survey analysis, there remains significant opportunity for further work. The expansion of this work can contribute meaningfully tinterdisciplinary collaboration between technology and healthcare sectors. Future improvements can include:

- Multilingual Analysis: Expanding sentiment analysis across regional languages can provide insights into non-English speaking populations.
- Real-Time Dashboards: Creating dashboards for healthcare providers to monitor spikes in mental health-related posts in real-time.
- Predictive Modeling: Applying machine learning to predict future emotional states or risk of mental health crises based on patterns in textual data.
- Collaboration with Healthcare Systems: Integrating findings with public health databases to enhance early intervention strategies.
- Inclusion of Audio/Video Data: Utilizing voice tone or video expressions (from platforms like YouTube) could offer richer mental health cues.

CONCLUSION

This project shows the power of combining social media and structured survey data to understand mental health trends. Real- time analysis helps identify emotional peaks, risky periods, and demographic segments most affected. These insights can assist policymakers, healthcare providers, and awareness campaigns to better design interventions. The combination of natural language processing and data visualization provides a scalable framework for ongoing mental health surveillance.

Future work could involve incorporating more advanced machine learning models for emotion classification, expanding data sources to include other languages and regions, and developing predictive models for crisis forecasting.

ACKNOWLEDGEMENT

I would like to express my sincere gratitude to Dr. Sunil Karimchandani for his valuable guidance, encouragement, and insightful feedback throughout the course of this research. His expertise and support were instrumental in shaping the direction and quality of this work.

REFERENCES

[1] S. Guntuku, M. Yaden, M. Kern, D. Ungar, and L. Eichstaedt, "Detecting depression and mental illness on social media: An integrative review," Current Opinion in Behavioral Sciences, vol. 18, pp. 43–49, June 2017.

[2] C. Resnik, W. Armstrong, and T. Claudino, "Using topic modeling to improve prediction of neuroticism and depression in college students," in Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 1348–1353, Lisbon, Portugal, Sep. 2015.

[3] M. De Choudhury, S. Counts, and E. Horvitz, "Predicting postpartum changes in emotion and behavior via social media," in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 3267–3276, ACM, Apr. 2013.

[4] M. Saha, D. Weber, "A Social Media-Based Examination of the Effects of Counseling Recommendations After Student Deaths on College Campuses," Journal of American College Health, vol. 67, no. 3, pp. 247–256, 2019.

[5] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," arXiv preprint arXiv:1301.3781, Jan. 2013.