Mental Health Analysis

Instructor- Dr. Ritu Arora

Cole Holub, Laxmi Narayanan Appu Saravanan, Safiuzzama Khan and Altamish Merchant

Department of Computer Science, College of Sciences, University of Texas at San Antonio,
San Antonio, TX, USA.

Abstract: *To analyze the mental health of the user by their social media posts in twitter, Instagram, reddit, Facebook and other platforms. Social media platforms have gained immense popularity in the past decade and have active participants which share most of their daily activities on these platforms. Many of the platforms also share much of their personal information along with the data open to the public for research and other purposes. social media is one of the largest sources of unstructured data. Analyzing that data and harvesting meaning out of that is a tedious job. Recently opinion mining has become an emerging topic due to the vast amount of opinioned data available on the various social networking sites. Microblogging has appeared relatively recently, and twitter is the most popular microblogging site used by the people. It is one of the biggest free, open data sources. Datasets from different social media platforms are helpful in many fields like sociology and psychology. In this work, the number of tweets in which opinions are highly unstructured and are either positive or negative or sometimes neutral is being used for the analysis. Sentiment Analysis can be supported with several machine learning algorithms to classify the mental health of the user. Simulation results show the enhancement in sentiment analysis of the Twitter data for the intuitive mental wellness of a person. This paper highlights the methodology for mental health analysis for the intuitive well-being of the person. It uses sentiment analysis and its approach to analyze the mental well-being of users and display results using data visualization.*

Keywords: Mental health, data, sentiment analysis, data visualization

1. INTRODUCTION

1.1 About

This project is focused on mental health and aims to understand the correlation between various factors and mental wellness. To gather information, data was collected from different sources such as social media platforms like Twitter, Instagram, and Facebook, as well as Pubmed library of articles and other websites related to mental health. The collected data was then analyzed and visualized using a dashboard on a website, which displays trends and statistics related to mental health. In addition to analyzing correlations between mental health and factors like food-related photos, the project also conducted sentiment analysis on social media data to identify individuals who may be struggling with stress, depression, or other mental health issues. Overall, this project provides insights into mental health and highlights areas where intervention may be needed to improve mental wellness.

1.2 Importance

The importance of this project lies in its ability to provide valuable insights into mental health and identify areas where intervention may be needed. Mental health issues are a significant concern worldwide, and there is a need for data-driven approaches to better understand these issues and develop effective interventions. By collecting and analyzing data from various sources, this project provides a comprehensive view of mental health and the factors that may influence it. This information can be used to inform policy decisions, develop new interventions, and improve existing programs.

Furthermore, the project's use of sentiment analysis on social media data is especially relevant, as social media platforms have become an important tool for communication and self-expression. Identifying individuals who may be struggling with mental health issues on these platforms could lead to early interventions that could improve their well-being.

1.3 Target audience

The target audience for this project is broad and includes anyone interested in mental health and wellness. This could include policymakers, mental health professionals, researchers, and members of the general public who are interested in improving their

mental health and well-being. Specifically, policymakers could benefit from the insights generated by this project to inform policy decisions related to mental health. Mental health professionals and researchers could use the data and analysis to further their understanding of mental health issues and develop new interventions. Additionally, members of the general public could use the insights generated by this project to make informed decisions about their mental health and well-being. They could also use the dashboard and search functionality on the website to access relevant information about mental health and identify areas where they may need to seek help.

2. DESIGN

The design of this project is focused on collecting and analyzing data from various sources related to mental health, including social media platforms like Twitter, Instagram, and Facebook, as well as PubMed library of articles and other websites related to mental health. The collected data is then analyzed using statistical and machine learning techniques to identify correlations between mental health and various factors such as food-related photos, sentiment analysis on social media data, and other relevant variables. To provide a comprehensive view of mental health, the project will use a dashboard on a website to display trends and statistics related to mental health. The dashboard will be user-friendly and interactive, allowing users to explore the data and analyze the trends in various ways. It will also include search functionality to access relevant information about mental health and identify areas where users may need to seek help. The design of this project is also intended to display the data of individuals who may be struggling with mental health issues on social media platforms. This will be accomplished by conducting sentiment analysis on social media data, which can help identify individuals who may be experiencing stress, depression, or other mental health issues. The project will use this information to develop interventions that could improve their well-being.

Overall, the design of this project is data-driven and aims to provide valuable insights into mental health. It is intended to be useful for a broad audience, including policymakers, mental health professionals, researchers, and members of the general public who are interested in improving their mental health and well-being.
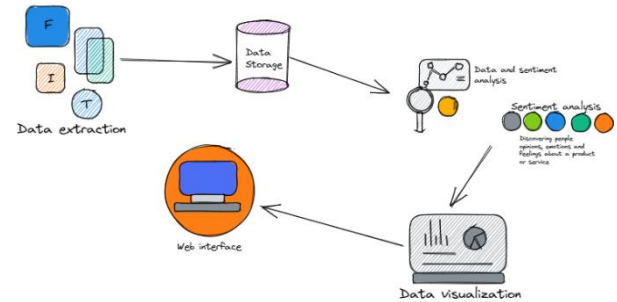

Fig1: Mental health analysis system design

3. IMPLEMENTATION

The implementation of this mental health project involves,

1. Data Collection: Collecting and extracting data from various sources, including social media platforms like Twitter, Instagram, and Facebook, as well as the PubMed library of articles and other websites related to mental health. Data is extracted using webscrapping scripts using python language where it extracts data and loads the data in format of .csv files. For scrapping data we use the topics and keywords that are related to mental health state like content related to food how people emotes their mental state from that so like this we have chosen the keywords like vacation, depressed, stressed, blissful, outing, travel, board, alone, loneliness, happy, sad and joyful. Our scrapper and by using the available APIs we collected large amount of data based on these keywords.

2. Data Cleaning: As data are collected in export files, using data cleaning method we filter the data and pre -process data for the next step. As part of this process we remove the quotes(" ") from the content text column in the data files we collected and generate new output files as a cleaned data for each keyword we chose to perform the sentiment analysis.

3. Data Storage: The pre-processed data from the cleaning is to be stored in the database. Using PostgreSQL server, created a database and imported the data from .csv files to the tables in database that are created based on the keywords we chose as part of sentiment analysis. On the tables created we have primary key column on the ID and created indexes for the tables on the column content which will have text data that is collected from the various sources. To maintain data integrity and quality all the data from various sources are merged together having same format and same column header for easy processing of data in the next stage. Amount of data stored is 8.54GB

in the database and images collection is about 4GB of size.

4.   Web UI: As a part of project we have created a web-interface and added the functionality of searching the database with the keywords to display the top 100 results of topics/keywords that are related to mental health. This search functionality allows people to search with emojis like heart, smileys related to emotions of people that describe about the mental health state. For this integration a PHP file is created that pings to the database server and searches for the keyword that is queried in webui in all the tables within the database and returns top 100 results of the content. Also added a dashboard to display trends/statistics on a website in form of result that are obtained from the sentiment analysis.

5.   Sentiment Analysis: We have done analysis based on keywords related to mental health for multiple topics {"vacation", "depressed", "stressed", "blissful", "outing", "travel", "board", "alone", "loneliness", "happy", "sad", "joyful", "food", "snack"} each word we are analyzing text content. c.   Analyze   the correlation of different factors such as photos and content that people tweet or share on social-media (e.g., photos that are related to vacation) and their state of health. Classifying them according to sentiment analysis and some common words where people mental state or mood depends on it based on these, we have collected data from social media and multiple sources and performed sentimental analysis based on this. This classification is based on polarity and subjectivity. Polarity includes the three traits positive, negative and neutral. We use this factor to do the main analysis of intuitive mental health. The higher the percentage of negative polarity, the higher the person's chances of having some mental health issue. Whereas the more positivity shown in their content in social media depicts their mental well-being. The analysis is then done for the mental health well-being on the basis of the polarity and subjectivity generated. Where high polarity indicates better mental health well-being and vice-versa. The analysis results are exported to the .csv files where the mood column is created and within that column it states whether the content share is positive/negative or neutral is being stored. Then this data is again reloaded to the database and tables with the mood column stating the mental health state of the people who expressed in the social media platforms. Also this result is being displayed in the webui search result part to show that content share is whether positive/negative or neutral as depicts the mental health of the people as these are flagged to indicate that a person may be stressed or may be needing help. This helps in identifying individuals who need help and support..

6.   Data Visualization: The results of our analysis have been displayed with appropriate methods of data representation such as graphs and word clouds. For each topics/keywords that are chosen are displayed in each graph where x-axis tells the result positive, negative and neutral where y-axis tells the count of people as sentiment analysis performed to depict their mental health state whether a positive emotion or neutral emotion or negative. To summarize overall topics/keywords related to mental health with the content shared by the people in the social media platforms we created the word clouds based on this to show the representation of the data.

7.   Security : To ensure the safety of data and server that is hosting the webpage contents we implement the IP proxy for the website and DDoS protection through Cloudflare. All our outside facing servers have login as root disabled and have randomly generated, 20-character passwords. We also have a very strict, minimal allowance firewall. And the permissions for our working accounts on the database and webserver are minimal privileges. Our SSL/TLS implementation is "A" rated by SSL labs using pfSense, let's encrypt and HAProxy.

Overall, the importance of this project lies in its ability to provide valuable insights into mental health and identify areas where intervention may be needed. The insights generated by this project can be used to inform policy decisions, develop new interventions, and improve existing programs.

4.   TESTING

The following are the testing procedures that will be used:

1.   Unit testing: Unit testing will be performed to ensure that each individual component of the project is functioning correctly. This will involve testing the code for the data collection process, data analysis, and visualization dashboard to ensure they work as expected. For data collections process we ensured web scraping script returns the expected data only. We validated results by ensuring the same amount of data is being imported and exported. Also, the ensured the content is of data being matched during the load. Also ensured that the features added in web user interface that is deployed and is working as expected.

2.   Integration testing: Integration testing will be conducted to ensure that all the individual components of the project are working together correctly. This will involve testing the interaction between the data collection, data storage, and dashboard visualization components. As part of this we have tested the server is being connected and stable between the database and

webserver that is hosting the webpage and returns the expected results for the search functionality.

3. User acceptance testing: User acceptance testing will be performed to ensure that the dashboard and website are user-friendly and meet the requirements of the target audience. While a search keyword is given in webui whether it is returning the accurate results as expected are not is also part of this testing and this has been validated. This will involve getting feedback from end-users and incorporating any necessary changes to improve the user experience in webui.

4. Stress testing: Stress testing will be conducted to ensure that the system can handle a large volume of data and traffic without crashing or slowing down. This will involve simulating high traffic and data volume scenarios to test the system's performance under heavy load.

5. Security testing: Security testing will be performed to ensure that the system is secure from cyber threats and that user data is protected. This will involve testing the system for vulnerabilities, implementing security measures, and conducting penetration testing to ensure that the system is secure. We use the personalized server for the database and to host the webserver. By performing these testing procedures, we can ensure that the mental health project is accurate, effective, user-friendly, and secure. Any issues identified during testing will be addressed promptly to ensure that the final product meets the highest standards.

5. PERFORMANCE

To ensure the accuracy and effectiveness of the mental health project, testing will be conducted at various stages of the implementation process. TBy collecting and analyzing data from various sources such as social media platforms and Pubmed library of articles, the project provides a comprehensive view of mental health and the factors that may influence it. The use of a dashboard on a website to display trends and statistics related to mental health also enhances the project's performance by making it easier to visualize the data and understand the insights generated. Additionally, the project's use of sentiment analysis on social media data is particularly relevant as it allows for the identification of individuals who may be struggling with mental health issues, enabling early interventions that could improve their well-being. The importance of this project in providing valuable insights into mental health and identifying areas where intervention may be needed further highlights its performance. Policymakers, mental health professionals, researchers, and members of the general public can use the insights generated by the project to inform policy decisions, further their understanding of mental health issues, develop new interventions, and make informed decisions about their mental health and well-being. Overall, the performance of this project is crucial in providing data-driven approaches to better understand mental health issues, develop effective interventions, and improve existing programs.

6. CHALLENGES

During the Mental Health Analysis project, we faced several challenges that required careful planning and problem-solving to overcome. One of the primary challenges we encountered was data collection. It was challenging to find a large dataset of social media posts related to mental health that was both relevant and representative of the community we were studying. We had to search extensively to find suitable data sources and ensure that the data was anonymized and compliant with ethical standards. Once we had collected the data, we faced another challenge: data cleaning and preprocessing. Social media data can be messy and unstructured, making it difficult to analyze. We had to spend a significant amount of time cleaning up the data, standardizing the language, and removing irrelevant content to prepare it for analysis. The accuracy of the sentiment analysis algorithm was another challenge we encountered. Sentiment analysis is not always accurate, especially when dealing with complex emotions related to mental health. We had to validate and test the accuracy of the algorithm extensively to ensure that our results were reliable. Bias and ethical concerns were also a significant challenge. There was a risk of bias in the data collected and analyzed, which could affect the accuracy of the results. We had to carefully consider the ethical implications of using social media data for mental health analysis, especially regarding data protection and privacy. Finally, interpreting the results of the analysis was challenging due to the complexity of mental health conditions and the multiple factors that contribute to mental health outcomes. We had to have a thorough understanding of the context and limitations of the analysis to ensure that the results were properly interpreted and applied. Overall, these challenges required careful planning, collaboration, and problem-

solving to overcome. By addressing these challenges, we were able to develop a robust methodology for analyzing mental health trends in a specific community, providing insights that could inform targeted interventions and resources to address mental health concerns.

7. RESULTS

Our project displays a system for the recognition of any issues related to mental health like a person sharing negative tweets is stressed out and could be dealing with depression and may be eating a lot of high-calorie food to feel better and may need help. Here we took results based on these topics { "vacation", "depressed", "stressed", "blissful", "outing", "travel", "board", "alone", "loneliness", "happy", "sad", "joyful", "food", "snack"} and performed analysis based on the correlation factors such as content related to the above mentioned keywords and mental health state. Based on these the sentiment analysis is performed to display the results in positive, negative or neutral and displayed it using data visualization in form of bar graphs and word clouds.

Sad Mental Health

Loneliness State

Alone Mental Health

Depressed Mental Health

Blissful Mental Health

Based on Food

Bored Mental Health

Happy Mental Health

Fig2: Data Visualization in bar graphs

Stressed Mental State

Mental Health based on outing

Mental Health based on snack

Joyful Mental Health

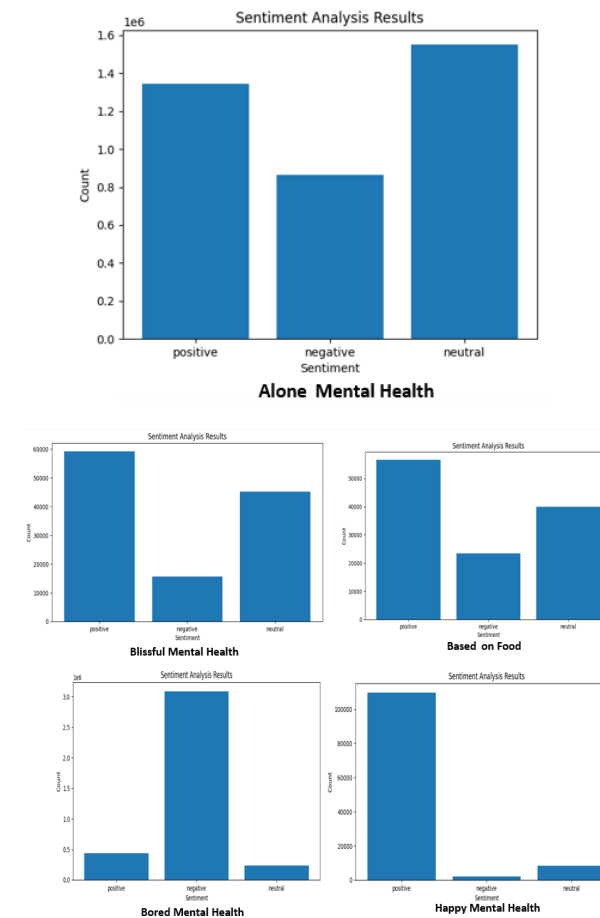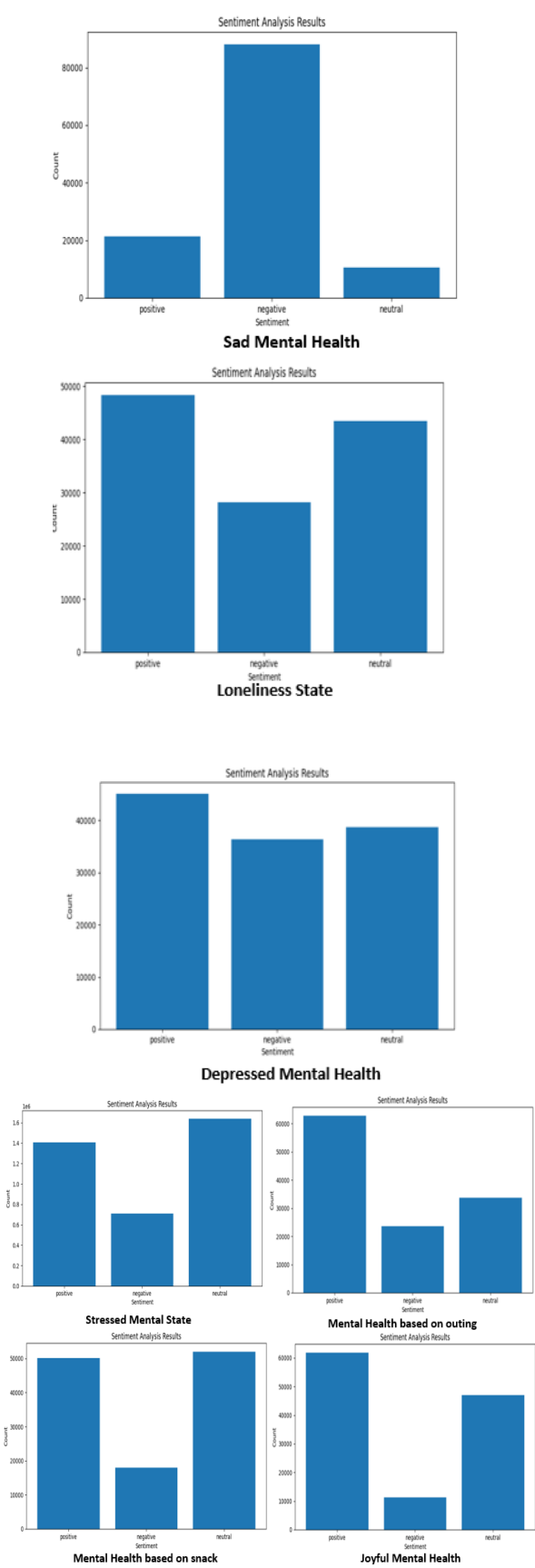*Fig3: Data Visualization in bar graphs*

5

Fig4: Data Visualization in wordcloud

8. REFERENCES

[1] Preeti Gupta , Mansi Gupta, Riya Kansal, Lovely Singh, Mahesh Khotani and VS Gupta. "Analysis of Twitter data to Identify Intuitive Mental Health Well Being". In Oct, 2021 5th International Conference on Information Systems and Computer Networks (ISCON).

[2] Aagam Shah, Rohan Shah, Praneeta Desai, Chirag Desai "Mental Health Monitoring using Sentiment Analysis" in International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056 Volume: 07 Issue: 07 | July 2020

[3] Anastazia Zunic, Padraig Corcoran, and Irena Spasic. "Sentiment analysis in health and wellbeing: A systematic review". In August 2019, JMIR Medical Informatics.