

Project Plan

Mental Health Analysis

Large Scale Data Management (CS-5493)

Objective	1
Deliverable	1
Team	2
Project plan	2
Project Planning	2
Data Collection	2
Data Cleaning	2
Data Storage	2
Web UI	2
Sentiment Analysis	2
Data Visualization	3
Security	3
Testing	3
Design draft	4
Milestones	4
References	5

Objective

The objective of the Mental Health Analysis project is to develop a methodology for analyzing social media posts to assess the mental health and well-being of users within a specific community, and to identify potential factors contributing to mental health conditions.

Deliverable

The project deliverables for the Mental Health Analysis project include a report detailing the methodology and findings, a presentation of the results, and a sentiment analysis tool developed for mental health-related social media posts. The report will provide insights into mental health trends in a specific community, along with recommendations for targeted interventions and resources. The sentiment analysis tool can be used to monitor and analyze mental health-related social media posts in real time, providing valuable information for mental health professionals, policymakers, and other stakeholders.

## Team

1. Cole Holub
2. Laxmi Narayana
3. Safiuzzama Khan
4. Altamish Merchant

## Project plan

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 web scrapping scripts using Python language where it extracts data and loads the data in the format of .csv files. For scrapping data, we use the topics and keywords that are related to mental health states like content related to food and how people emote 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 a large amount of data based on these keywords.
2. **Data Cleaning:** As data are collected in export files, using the 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 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 the PostgreSQL server, created a database and imported the data from .csv files to the tables in the database that are created based on the keywords we chose as part of sentiment analysis. On the tables created, we have a 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 having the same format and same column header for easy processing of data in the next stage. The amount of data stored is 8.54GB in the database and image collection is about 4GB in size.
4. **Web UI:** As a part of the 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 hearts, and smileys related to emotions of people that describe their 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 Web-UI in all the tables within the database and returns the top 100 results of the content. Also added a dashboard to display trends/statistics on a website in the form of results 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 between different factors such as photos and content

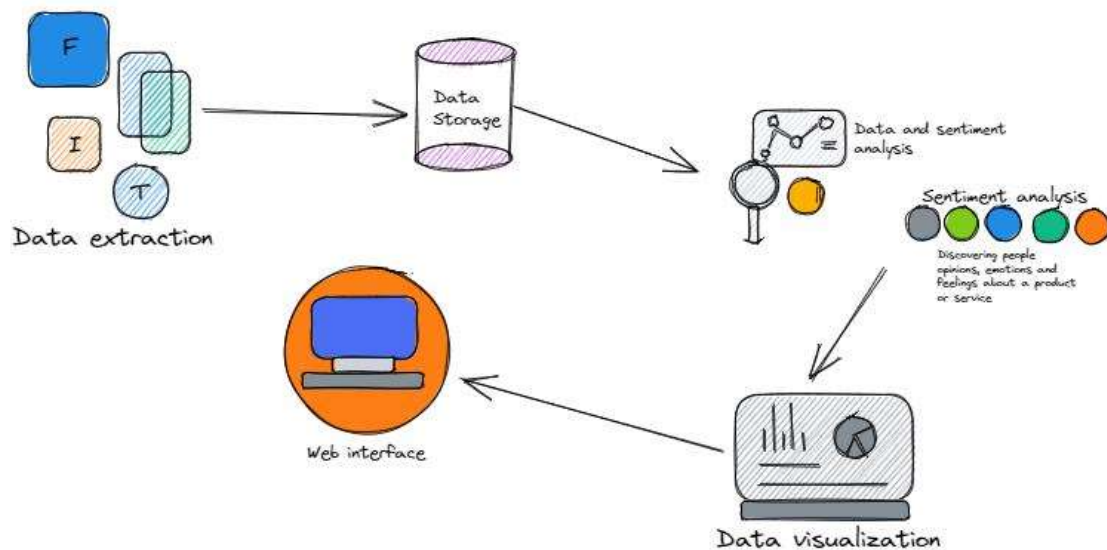
that people tweet or share on social media (e.g., photos that are related to vacation) and their state of health. Classifying them according to sentiment analysis and some common words where people's mental state or mood depends on it based on these, we have collected data from social media and multiple sources and performed a sentimental analysis based on this. This classification is based on polarity and subjectivity. Polarity includes 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 on social media depicts their mental well-being. The analysis is then done for mental health well-being based on 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 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 topic/keyword that is chosen is displayed in each graph where the x-axis tells the result positive, negative, and neutral where y-axis tells the count of people and sentiment analysis is 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 the 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 web server are minimal privileges. Our SSL/TLS implementation is "A" rated by SSL labs using pfSense.
8. **Testing:** The testing procedures for the mental health project include unit testing, integration testing, user acceptance testing, stress testing, and security testing. Unit testing ensures each component functions correctly while integration testing ensures all components work together. User acceptance testing evaluates user-friendliness, stress testing simulates high traffic, and security testing ensures protection from cyber threats. Any issues will be promptly addressed to ensure a high-quality final product.

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.

## Design draft

The above plans is depicted in the diagram below:



## Milestones

Task	Timeline	Task owner
Identify project goals, scope, and deliverables	3/6/2023	Everyone
Conduct research on relevant technologies and tools	3/8/2023	Everyone
Develop a detailed project plan	3/13/2023	Everyone
Collecting data on your selected topics from a variety of sources like Twitter, Instagram, Facebook, PubMed library of articles, and other websites related to Mental Health.	3/15/2023	Cole Holub, Altamish Merchant, Laxmi Narayanan
Writing web-scraping scripts, using the available API for data collection, downloading datasets directly if that is supported by the source of data, and/or reaching out to the data owners with requests for a copy of the data.	3/20/2023	Cole Holub, Altamish Merchant
Adding a dashboard to display trends/statistics on a website and adding the functionality of searching the database	3/22/2023	Laxmi Narayanan
Analyze the correlation between different factors such as photos that people tweet or share on social media (e.g., photos that are related to food) and their state of health	3/27/2023	Safiuzzama Khan

Do sentiment analysis on the social media data and check the correlation between the sentiments and wellness of an individual	3/29/2023	Cole Holub, Altamish Merchant
Flag the tweets that indicate that a person may be stressed may be needing help with mental health	4/3/2023	Safiuzzama Khan, Laxmi Narayanan
Display the results of the selected analysis with appropriate methods of data representation using graphs, tables, and heat maps.	4/5/2023	Cole Holub, Altamish Merchant
Evaluate the model performance using appropriate metrics	4/10/2023	Altamish Merchant
Compare the model's performance with existing methods	4/12/2023	Laxmi Narayanan
Report the results and discuss the implications	4/17/2023	Safiuzzama Khan
Prepare documentation of the project, including the project plan, code, and results	4/27/2023	Safiuzzama Khan, Laxmi Narayanan
Prepare a presentation of the project and present it	5/1/2023	Laxmi Narayanan
Submit the final project report and deliverables	5/1/2023	Everyone

## 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.