**Depression: Twitter Dataset + Feature Extraction**

<https://www.kaggle.com/datasets/infamouscoder/mental-health-social-media?resource=download>

The data is in uncleaned format and is collected using Twitter API. The Tweets has been filtered to keep only the English context. It targets mental health classification of the user at Tweet-level. Also check out notebooks I have provided which demonstrates Data Cleaning and Feature Extraction Techniques on the given dataset

1. **Topic Modelling Features** using LDA (Latent Dirichlet Allocation) i.e. summarizing tweet into one of Top k topics
2. **Emoji Sentiment Features** i.e. count of Positive, Negative and Neutral Expression emoji's present in the tweet

Reddit Mental Health Data

<https://www.kaggle.com/datasets/neelghoshal/reddit-mental-health-data>

Sourced from subreddits concerned with mental health based conversations  
Targets given have the following mappings:  
0 = Stress  
1 = Depression  
2 = Bipolar disorder  
3 = Personality disorder  
4 = Anxiety

**Sentiment Analysis for Mental Health**

**About Dataset**

**This comprehensive dataset is a meticulously curated collection of mental health statuses tagged from various statements. The dataset amalgamates raw data from multiple sources, cleaned and compiled to create a robust resource for developing chatbots and performing sentiment analysis.**

**Data Source:**

**The dataset integrates information from the following Kaggle datasets:**

* [**3k Conversations Dataset for Chatbot**](https://www.kaggle.com/datasets/kreeshrajani/3k-conversations-dataset-for-chatbot)
* [**Depression Reddit Cleaned**](https://www.kaggle.com/datasets/infamouscoder/depression-reddit-cleaned)
* [**Human Stress Prediction**](https://www.kaggle.com/datasets/kreeshrajani/human-stress-prediction)
* [**Predicting Anxiety in Mental Health Data**](https://www.kaggle.com/datasets/michellevp/predicting-anxiety-in-mental-health-data)
* [**Mental Health Dataset Bipolar**](https://www.kaggle.com/datasets/michellevp/mental-health-dataset-bipolar)
* [**Reddit Mental Health Data**](https://www.kaggle.com/datasets/neelghoshal/reddit-mental-health-data)
* [**Students Anxiety and Depression Dataset**](https://www.kaggle.com/datasets/sahasourav17/students-anxiety-and-depression-dataset)
* [**Suicidal Mental Health Dataset**](https://www.kaggle.com/datasets/aradhakkandhari/suicidal-mental-health-dataset)
* [**Suicidal Tweet Detection Dataset**](https://www.kaggle.com/datasets/aunanya875/suicidal-tweet-detection-dataset)

**Data Overview:**

**The dataset consists of statements tagged with one of the following seven mental health statuses:**

* **Normal**
* **Depression**
* **Suicidal**
* **Anxiety**
* **Stress**
* **Bi-Polar**
* **Personality Disorder**

**Data Collection:**

**The data is sourced from diverse platforms including social media posts, Reddit posts, Twitter posts, and more. Each entry is tagged with a specific mental health status, making it an invaluable asset for:**

* **Developing intelligent mental health chatbots.**
* **Performing in-depth sentiment analysis.**
* **Research and studies related to mental health trends.**

**Features:**

* **unique\_id: A unique identifier for each entry.**
* **Statement: The textual data or post.**
* **Mental Health Status: The tagged mental health status of the statement.**

**Usage:**

**This dataset is ideal for training machine learning models aimed at understanding and predicting mental health conditions based on textual data. It can be used in various applications such as:**

* **Chatbot development for mental health support.**
* **Sentiment analysis to gauge mental health trends.**
* **Academic research on mental health patterns.**

**Acknowledgments:**

**This dataset was created by aggregating and cleaning data from various publicly available datasets on Kaggle. Special thanks to the original dataset creators for their contributions.**

<https://www.kaggle.com/datasets/suchintikasarkar/sentiment-analysis-for-mental-health>

Detection and prediction of Future Mental disorder from Social Media Data using Machine Learning, Ensemble Learning, and Large Language Models.

<https://www.researchgate.net/publication/380952538_Detection_and_prediction_of_Future_Mental_disorder_from_Social_Media_Data_using_Machine_Learning_Ensemble_Learning_and_Large_Language_Models>

Predicting mental health using social media: A roadmap for future development

<https://www.researchgate.net/publication/367432297_Predicting_mental_health_using_social_media_A_roadmap_for_future_development>

Diagnosis of mental disorders using machine learning: Literature review and bibliometric mapping from 2012 to 2023

<https://www.sciencedirect.com/science/article/pii/S2405844024085797>

**Mental Health Prediction Using Machine Learning: Taxonomy, Applications, and Challenges**

<https://onlinelibrary.wiley.com/doi/10.1155/2022/9970363>