Useful source:

<https://www.google.com/search?q=mark+dredze&oq=mark+dredze&gs_lcrp=EgZjaHJvbWUyBggAEEUYOTIHCAEQABiABDIHCAIQABiABDIICAMQABgWGB4yCggEEAAYDxgWGB4yCggFEAAYDxgWGB7SAQkxNTY2NmoxajGoAgCwAgA&sourceid=chrome&ie=UTF-8>

<https://www.google.com/search?q=munmun+de+choudhury+google+scholar&oq=munmun+de+&gs_lcrp=EgZjaHJvbWUqBwgCEAAYgAQyBggAEEUYOTIHCAEQLhiABDIHCAIQABiABDIICAMQABgWGB4yBggEEAUYLNIBCDg1NDZqMGoxqAIAsAIA&sourceid=chrome&ie=UTF-8>

<https://www.google.com/search?q=resnick+maryland&oq=resnick+maryland&gs_lcrp=EgZjaHJvbWUyBggAEEUYOTIICAEQABgWGB4yCAgCEAAYFhgeMggIAxAAGBYYHjIICAQQABgWGB7SAQg5NDY5ajBqMagCALACAA&sourceid=chrome&ie=UTF-8>

<https://perspectiveapi.com/>

<https://github.com/kharrigian/mental-health-datasets>

<https://tweetsets.library.gwu.edu/datasets>

What is the question and who is the audience? Who will care about finding out the answer

What is the dataset that will help you to answer the question

<http://users.umiacs.umd.edu/~resnik/umd_reddit_suicidality_dataset.html>

<https://catalog.data.gov/dataset/?tags=mental-health>

Is bigger data better? Where has the data gone? (twitter api gone, cambridge analytics)

Are we able to answer the question? If the data is correct? Different types of questions

NLP for measuring an individual's mental health?

**Can mental health disorders be predicted by social media activities/behavior?**

<https://www.nature.com/articles/s41746-022-00589-7>

<https://mental.jmir.org/2022/12/e39747/>

<https://aclanthology.org/D19-5542/>

<https://link.springer.com/article/10.1007/s10796-022-10282-5>

<https://ojs.aaai.org/index.php/ICWSM/article/view/14526>

Dataset- reddit self reported mental illness

<https://ir.cs.georgetown.edu/resources/rsdd.html>

Dataset - twitter api mental illness

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

Reddit - mental health dataset (28 subreddit)

<https://zenodo.org/record/3941387>

<https://zenodo.org/record/4266616>

chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/<https://scholarworks.gsu.edu/cgi/viewcontent.cgi?article=1186&context=cs_diss>

Basic idea for the project:

* How do we measure mental health disorder?
* What social media platform are we using?
* Using NLP for individual activities and behavior
* Classification that we are using: svm
* Types of mental illness

Begin with broad questions narrow down, focus in operationalize observe analyze data reach conclusions generalize back to questions

Types of questions

* Inductive: practical problems in the field
* Deductive: literature/prior theory
* Abductive: can i apply x to y?
* Directive from sponsor or request for proposal
* Personal interest
* Descriptive: what relationships exist
* Relational: does this specific relationship exist?
* Causal: is the observed, existing relationship caused by something else?

Does it address the problem?

Is it too broad or too specific?- who will care about the answer, is it feasible

What happens if your results don’t come out as anticipated?

Related/Useful papers:

chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/<https://aclanthology.org/W14-3207.pdf>

<https://ieeexplore.ieee.org/document/9230410>

<https://academic.oup.com/jamia/article/24/3/496/2907899#google_vignette>

<https://ojs.aaai.org/index.php/ICWSM/article/view/14526/14375>

https://www.sciencedirect.com/science/article/pii/S2667096822000568