1. **Who cares?**

Who will care? •What are the implications for the customer, and for society, if you answer this question or solve this problem?

Individuals - behavior, actions, activities/ society

Mental health related subreddits

NLP for measuring an individual's mental health?

Scaled depression level, clustered keywords

**Happiness index - hedonometrics**

**Individuals - for their awareness, happiness; people around etc**

**Professionals - Healthcare (mental health-specialized) clinicians, researchers -> early signs, diagnosis, treatment**

**Society - how socio-economic circumstances of individuals likely to affect the development of mental disorder**

1. **3 version of questions**

We need to know how social media reflects mental illness. We want to use several social media to track people’s behavior on social media. Counts words

How social media can predict

* What social media platform are we using? Reddit, twitter, instagram if we can
* Using NLP, types of mental illness - depression, bipolar, anxiety, adhd
* ML Classification that we are using: svm, k nearest, naive bayes (??)

**Inductive: Are emotions and behaviors one expressed on social media a reflection of their mental health? Start with the data, how data related to observations**

**From negative words on reddit threads data, how it measure individual’s mental illness?**

**Deductive: Is mental illness predictable from behaviors and activities on social media?**

**Start with the theory- when depressed expected to find sth**

**When people with depression post on reddit threads, are we able to collect their behavior from their posts??**

**Abductive: Given mental illness symptoms, does social media would be the most likely cause of an individual's mental disorder? / given social media keywords, does mental illness predictable from their interactions and behaviors**

**Can mental illness symptoms be measured by individual behavior on social media threads?**

Inductive: practical problems in the field, building theory based on data

Deductive: literature/prior theory, testing theory based on observation

Abductive: can i apply x to y?

Directive from sponsor or request for proposal

Personal interest

1. **Express question in 3 ways and justify**

**Descriptive:**

**Can verbal expressions on social media primarily examine the user’s mental health and reflect their behaviors?**

**Relational:**

**Is there any correlation between a user's mental illness such as depression and negative expressions on social media?**

**Causal: Does mental illness can cause negative verbal expression/behavior on social media?**

**Our choice is**

Given the goal and data source, descriptive

Descriptive: what relationships exist? - public poll

Relational: does this specific relationship exist? - relationship between gender and voting preference

Causal: is the observed, existing relationship caused by something else? - whether campaign cause changed the proportion of voters would vote demo or repub (effect)

Social Media is a new phenomenon that allows social interaction between people known/unknown using Internet-based applications

Justify, given your goal and data source, which type of question is best suited to your overall goal

•Descriptive: What relationships exist? •Relational: Does this specific relationship exist? •Causal: Is the observed, existing relationship caused by something else?

1. **What is your unit of analysis? Justify**

Individuals, groups, artifacts-books, photos newspapers, geographical units - state, census tract, social interactions-divorces, arrests

**Our unit of analysis is individual’s social media activities -> quantifying their emotions from threads (post and comments ) to measuring mental health**

1. **Is your question cross-sectional or longitudinal and why?**

**Cross-sectional - analyzing the positivity/negative level of a tweet/post at one point in time and draw conclusion of the individual’s mental state in that specific point**

**subreddits on mental illness -> individual depression by**

Time is not a factor

1. **Goldilocks**

- Too broad: What are the effects of using social media on individuals’ health?

- Too specific: What is the percentage of tweets that indicates negative sentiments in the US?

- Just right: Can mental health disorders be predicted by social media activities/behavior?

- Justification: This question adequately points out which data will be collected and analyzed and which conclusion to be made after that. It also calls for an appropriate discussion if needed.

1. **Feasibility**

•How long will it take? • Is it ethical? •Who is on the critical path? •How much will it cost? • Is it appropriately scoped?

**How long: Need to gather the dataset relevant to mental illness, build models**

**Currently, we need to gather the dataset regarding to mental illness reddit threads, and after that we need to build a model based on what we collected**

**Ethical: (It is ethical) Debatable since we need access to individuals' mental health records and will present it in class, will consider a different route if this data cannot be obtained or if we deem it too personal and cannot be anonymized**

**Critical path: data collection (from reddit, twitter if we can find it) → data preprocessing (data cleaning for keywords, clean up unnecessary words) → natural language process (build model on mental illness) → classification (ml for which word represents which emotion, what kind of word related to specific mental disorder) → visualization and conclusion**

**Reaching out to georgetown.. Critical path.. Who is on the critical path and what is the problem? Under in the control**

**During the data collection, we reached out to georgetown - for Reddit self-reported depression diagnosis dataset. This dataset is the reddit posts of app. 9000 users who have claimed to have been diagnosed with depression and app. 107,000 matched control users. However, they are only allowed to share with research purposes so they are on our current critical path. If we cannot collect this dataset, we need to find an alternative dataset to measure an individual's mental illness.**

**four mental health related Reddit communities with the largest number of publications, namely Depression, SuicideWatch, Anxiety e Bipolar.**

**Reddit Mental Health Dataset**

**Contains posts and text features for the following timeframes from 28 mental health and non-mental health subreddits:**

* **15 specific mental health support groups (r/EDAnonymous, r/addiction, r/alcoholism, r/adhd, r/anxiety, r/autism, r/bipolarreddit, r/bpd, r/depression, r/healthanxiety, r/lonely, r/ptsd, r/schizophrenia, r/socialanxiety, and r/suicidewatch)**
* **2 broad mental health subreddits (r/mentalhealth, r/COVID19\_support)**
* **11 non-mental health subreddits (r/conspiracy, r/divorce, r/fitness, r/guns, r/jokes, r/legaladvice, r/meditation, r/parenting, r/personalfinance, r/relationships, r/teaching).**

**Cost: for our data collection, it does not cost any but we need the time for collecting the threads dataset regarding to mental illness**

**Appropriately scoped: scope is large- what do ppl post when they are depressed?? - too big.. Narrow scope - one particular subreddit ; right balance and justify why would they care?? Sake of argument**

**What is the scope and why is it the right scope**

**Our scope is in right balance that using particular subreddit regarding to mental health - depression, anxiety, bipolar, suicide. We aim to analyze their expression on social media to recognize their mental illness. As people with negative behavior are highly predictable to write or react negatively on social media as well. Also, reddit is the online social interaction where people can express their feelings without limitation**

1. **You can't lose?**

What if the answer to your question is “no”? What if it’s “yes”? – discuss how to formulate your question so that you “can’t lose”

**We expect the answer to be “yes" since humans are prone to use social media as an outlet for emotions because it's accessible, fast and easy. Also they receive instant gratification in the form of likes, comments, and shares**

**so any social media activities with negative undertones would likely indicate a bad state of mind, which can potentially act as an early sign of mental illness**

**If the answer is “no", we will reformulate the question to “Does mental illness can provide**

**Feedbacks:**

Legally: implication of obligation of the target customers

Consequences of false positives/negatives

Requirements of the dataset (training certificates etc…)

Mental illness → mental health

Ethics requirement

<https://authorservices.taylorandfrancis.com/editorial-policies/research-ethics-guidelines-for-stem-journals/>