**DSC500 Ex. 9.2**

**Makayla McKibben**

**07.30.2024**

**Introduction for my DSC520 Final Project for Context**

Global awareness and tolerance of mental health conditions have been increasing rapidly in the last few years. However, a recent comprehensive analysis of the prevalence and most effective treatments of anxiety is not something I've come across. Mental illness and, specifically, anxiety is a topic that affects a great many people. Whether you are afflicted or have a spouse, relative, or friend who deals with mental illness, the impact of anxiety on a person's quality of life can touch nearly everyone. In order to determine the prevalence of anxiety and the effectiveness of various treatments, we will look at several datasets that provide recent, relevant information. R will be critical in analyzing this much data from these expansive datasets. While the World Health Organization published a study in 2019, and the NIH published a study using data from 2001 - 2004, I'd like to see how a more recent analysis will compare with their findings(WHO, 2023, NIH, 2010). This data and our inferences can be checked against the 2019 WHO analysis to see the progression of anxiety's pervasiveness and treatment options in the last five years and against the NIH study to see the change in the last 20 years.

**What if we Changed the Approach? Machine or Deep Learning Empowered Analysis.**

Instead of me analyzing the datasets manually, what would be the implications of having an ML program look at the data and try to determine if someone A) Has anxiety, B) Needs medication, and C) What the best medication for them would be? We could potentially use machine learning to set the foundation for this decision-making. We'd need to feed the ML program a lot of data, and the datasets I've selected from Kaggle for my DSC520 final project would be a good start; they are reachable by following the url in this paper’s references section (Kaggle, 2024). In addition, while working on my DSC520 project, I found datasets irrelevant to my DSC520 project, but they would help feed the ML program. The additional datasets I'm thinking of for this involve text responses and social media posts from people with anxiety.

There are some substantial potential problems with using ML for this. What happens if the ML program makes the wrong decision is the most glaring to me. Does the person in question not receive care? I'd imagine the initial implementation of this ML program could be for using it as a screening tool. It could look for cases where it is definite that the person is exhibiting anxious behavior and flag them for review with the person's doctor but not exclude a person from receiving care if the ML program does not flag them. Flagging personal data leads us to another dilemma: How much of this person's private health information would the ML program need access to for the program to be successful? And how much of their digital information would the program would need to look through? What is considered too invasive? These considerations need to be reviewed, and even with review, we may not reach a unanimous conclusion as it is subjective. Person A may be okay with the program tracking their browser history, while person B says absolutely not.

Tackling part B, I'm unsure how we would teach the program to determine when medication is warranted; it seems very subjective and is case-by-case with an actual doctor, let alone trying to set rigid parameters for a program to make decisions from. For the sake of brevity, I will acknowledge that part B poses problems and requires a substantial amount of thought and move on.

Problem C may be the best situation for implementing a program of this sort. We can utilize ML or DL to look at a holistic picture of a person (demographics, economic status, stress factors, family history, previous medical history, etc.) and tell it which medication worked best for said person. In that case, we can predict which medication would work best for someone with a similar profile. This implementation of the program still poses problems. Will the algorithm make the same decision for people with similar profiles but with different circumstances? How does one quantify someone's outlook on life, which could be a factor in why they are anxious? Is it even ethical to consider using a machine or deep learning program this way?

This proposed implementation of M/DL is rife with problems, both ethical and practical, that require input from the community at large. I think that if we could reach a consensus on parameters to use for the algorithm, then some of the ethical issues could be alleviated with a voluntary waiver. While the subject of mental illness is important to me personally, as stated in my DSC520 introduction, it impacts nearly everyone in some way. Using technology in sensitive matters can be problematic and is made even more complicated when the matter has subjective components. However, companies are farming our information; why not harness it to help ourselves and each other?

References

U.S. Department of Health and Human Services. (2010). *Any anxiety disorder*. National Institute of Mental Health. https://www.nimh.nih.gov/health/statistics/any-anxiety-disorder

World Health Organization. (2023). *Anxiety disorders*. World Health Organization. https://www.who.int/news-room/fact-sheets/detail/anxiety-disorders

*Your home for data science*. Kaggle. (2024). https://www.kaggle.com/work/collections/14401980