Sprint 0: Sentiment Analysis for Mental Health

Author: Gennaro Costantino

Date: Aug 21, 2024

Project: Capstone DS Brainstation

The Problem Area:

My area of interest lies in understanding public sentiment surrounding mental health issues as expressed on social media platforms. In today's digital age, mental health is increasingly discussed online, where individuals often share their experiences, seek support, or express opinions on mental health topics. Despite the wealth of data available, there is a significant challenge in accurately gauging the overall sentiment of these discussions. The key problem I aim to address is how to automatically and accurately analyze these sentiments to understand the broader public perception of mental health issues. This could potentially highlight areas where more awareness or support is needed, and identify trends that could inform mental health advocacy and policy-making.

The User:

The primary users who would benefit from this project are mental health professionals, researchers, people (families and friends) and policymakers. These stakeholders often lack the tools to efficiently monitor and analyze public sentiment on mental health, which can be crucial in shaping effective communication strategies, public health campaigns, and policy decisions. By providing insights into how mental health is discussed online, this project could help them better understand the public's needs and concerns, ultimately leading to more targeted and impactful interventions.

The Big Idea:

Machine learning, particularly in the form of Natural Language Processing (NLP), can offer robust solutions for analyzing vast amounts of text data from social media platforms. Sentiment analysis models can be trained to automatically classify posts as positive, negative, or neutral, providing a real-time gauge of public sentiment. Moreover, topic modeling techniques can be used to identify the main themes in mental health discussions, such as depression, anxiety, or stigma, and understand how these topics evolve over time. Previous research in this area has utilized various machine learning approaches, including supervised learning models like Logistic Regression and deep learning models like BERT, to achieve high accuracy in sentiment classification. By building upon these techniques, this project aims to create a tool that not only classifies sentiment but also provides deeper insights into the context and drivers of public opinion on mental health.

The Impact:

The societal value of this project lies in its potential to improve mental health outcomes by enabling more informed decision-making among stakeholders. Understanding public sentiment can help mental health organizations and policymakers to tailor their messages more effectively, address misinformation, and respond more rapidly to emerging issues. For example, if the model detects a

spike in negative sentiment related to a particular mental health issue, interventions can be designed to address this, potentially preventing escalation. The economic impact could also be significant, as better-informed campaigns and policies could lead to more efficient allocation of resources, reducing the overall cost of mental health services by targeting them where they are most needed.

The Data:

Several datasets can be leveraged for this project:

- **Sentiment140 Dataset:** Contains 1.6 million tweets labeled as positive, negative, or neutral, and can be used to train sentiment analysis models.
- **Mental Health Tweets Dataset:** A specific dataset containing tweets related to mental health, which can be useful for fine-tuning the sentiment analysis model to focus on mental health topics.
- **Reddit Mental Health Dataset:** Contains posts from mental health-related subreddits, which can provide more in-depth discussions and context compared to Twitter data.
- **Google Trends Data:** Can be used to track the popularity of mental health-related search terms over time, providing an additional layer of analysis to complement social media data.

These datasets collectively offer a robust foundation for building and testing sentiment analysis models focused on mental health discussions.

The Alternative:

An alternative subject area that interests me is the impact of social media usage on mental health, particularly among adolescents. This project would focus on analyzing how different types of social media interactions (e.g., likes, comments, shares) correlate with mental health outcomes such as anxiety or depression, using a combination of survey data and social media activity logs.