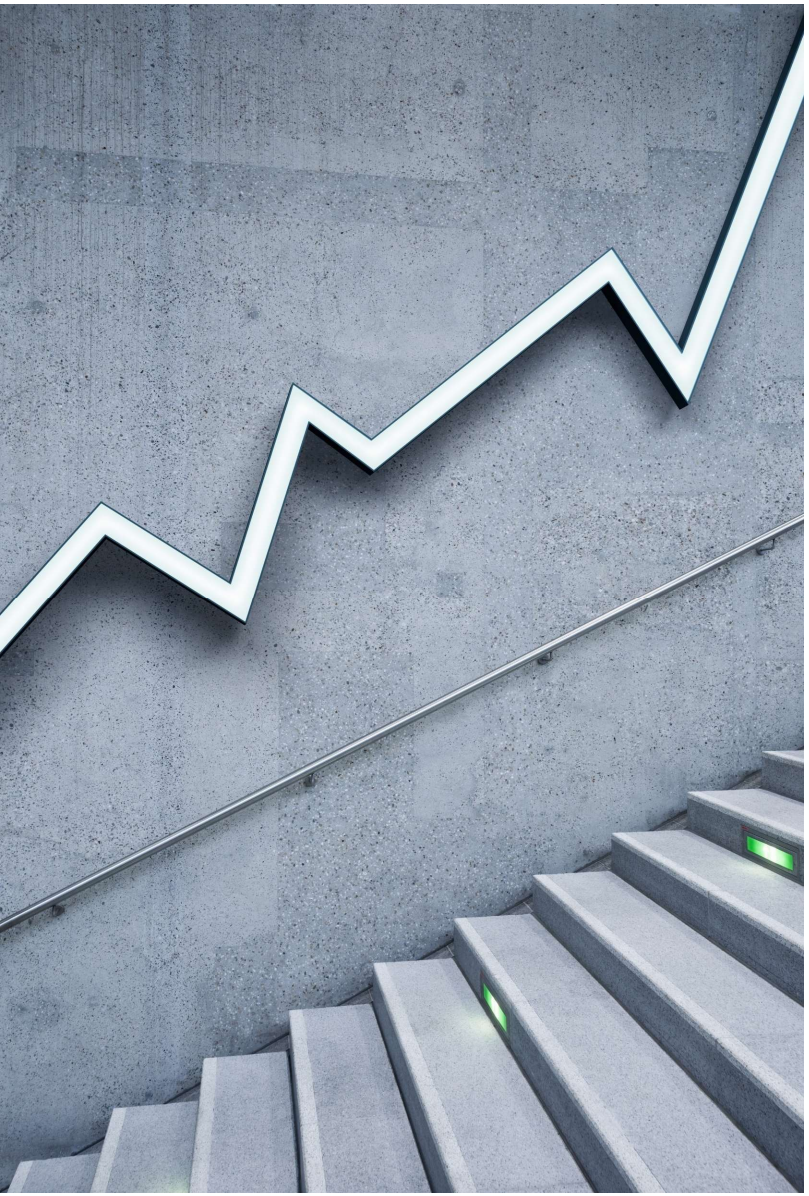


Sprint 1

Data Science Bootcamp

GC

Sentiment Analysis for Mental Health: A Data Science Approach

Non-Technical Overview: 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.

Problem Statement: Leveraging advanced machine learning and natural language processing (NLP) techniques to analyze vast amounts of textual data from social media and other platforms. This analysis can uncover insights into public sentiment around mental health issues, identify key topics of discussion, and track how these sentiments evolve over time.

Data Science Vision for Sentiment Analysis

Proposed Solution:

The ability to automate the analysis of sentiment in mental health-related conversations.

Approach Overview:

- Scalability: Sentiment analysis can process and analyze large datasets, making it possible to understand public sentiment on a massive scale—far beyond what manual analysis could achieve.
- Timeliness: By applying real-time sentiment analysis, organizations can respond quickly to shifts in public mood, enabling more agile and effective interventions.
- Contextual Understanding: Beyond simple positive or negative labels, sentiment analysis can provide a nuanced understanding of the public's feelings about specific mental health topics, such as depression, anxiety, or stigma.
- Data-Driven Insights: Sentiment analysis can transform unstructured text data into actionable insights, helping to inform decisions in mental health advocacy, public health campaigns, and policy development.

Potential Impact

- Real-time Monitoring: Develop tools that can monitor social media in real-time to detect shifts in public sentiment about mental health issues.
- Targeted Mental Health Campaigns: By understanding the emotional tone and key topics in public discourse, mental health organizations can tailor their communication strategies to address specific concerns, counteract stigma, and promote mental health awareness more effectively.
- Policy and Decision-Making: Policymakers can use insights from sentiment analysis to inform mental health policies that resonate with the public's needs and concerns.
- Support for Mental Health Professionals: Mental health professionals can benefit from understanding the broader social context in which their patients operate.

Dataset Overview & Exploratory Data Analysis (EDA)

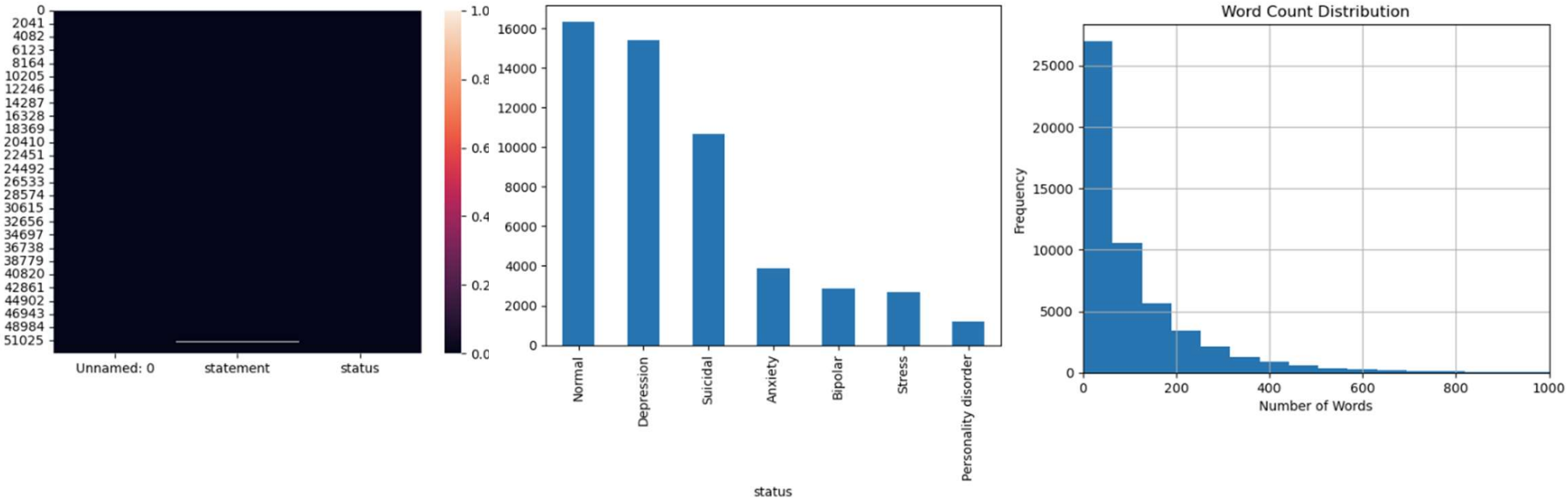
Datasets Introduced:

- Kaggle Datasets: Sentiment Analysis for Mental Health.
- 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.

Dataset Overview & Exploratory Data Analysis (EDA)

Preliminary EDA Findings using the Sentiment Analysis for Mental Health dataset (Kaggle):

The preliminary EDA of the "Sentiment Analysis for Mental Health" dataset highlights missing data in the statement column, a diverse distribution of mental health statuses, and a predominance of short text entries typical of social media.



Next Steps in Data Processing and Modeling

Data Processing:

- Text Preprocessing: Continue cleaning and preparing the text data (tokenization, stopword removal).
- Feature Engineering: Create features from the processed text data (e.g., TF-IDF vectors).

Baseline Modeling:

- Initial Model: Start with a baseline sentiment analysis model (e.g., Logistic Regression).
- Model Evaluation: Evaluate the model's performance using metrics like accuracy, precision, and recall.

Integration & Analysis:

- Merge Sentiment Analysis model to automatically classify posts as positive, negative, or neutral, providing a real-time gauge of public sentiment.
- Refinement: Based on initial findings, refine models and explore more advanced techniques if necessary.

Thank you

GC