# How the Disabled Community Talks About Disability: A Sentiment and Predictive Analysis of Reddit Comments

Tace Harris
Quantitative Methods in The Social Sciences, The University of Michigan
QMSS 301: Quantitative Social Science Analysis and Big Data
Dr. Korede Ajogbeje
April 21st, 2025

***Introduction/ Problem Statement***

Disability is a critical social issue that affects millions of individuals, yet the lived experiences of disabled people are often underrepresented or mischaracterized in public discourse. While much research focuses on how society perceives disability, it is equally important to understand how disabled individuals describe and frame their own experiences. This project draws on a Reddit thread titled "People with disabilities, what is something that's normal for others but a nightmare for you?" in which users with disabilities respond candidly to a widely shared prompt. The post provides a valuable corpus of first-person narratives, offering insight into how disabled people communicate the daily challenges, frustrations, and stigmas they face.

By analyzing thousands of comments using sentiment analysis (via VADER) and predictive modeling (via Scikit-learn), this research investigates the emotional tones, common themes, and features associated with positive or negative expressions. Rather than examining how others view disability, this project focuses on how people with disabilities narrate their own realities in a digital space. The central research problem is to determine how sentiment and textual characteristics in self-reported disability narratives can be systematically analyzed to reveal patterns in emotional expression and lived experience. In doing so, this project contributes to a richer understanding of the social and emotional dimensions of disability as expressed by those most directly affected.

***Data Collection***

To explore how people with disabilities describe their daily challenges and frustrations, I intend to analyze a Reddit post from the *r/AskReddit* subreddit titled *"People with disabilities, what is something that's normal for others but a nightmare for you?"* This post invites users who identify as disabled to candidly share their personal experiences. As of the time of collection, the thread contains 2,063 comments, making it a robust and relevant source of self-reported narratives centered on disability.

To collect the data I will use the Python Reddit API Wrapper (PRAW) to scrape the post and extract comment's text, score (upvotes), author ID, and UTC timestamp. To support further analysis, I will preprocess the comments by cleaning the text and removing filler words. I will also create two additional columns: one for the number of characters in each comment and another for the number of words. These features will help capture comment length and depth of expression, which may relate to sentiment and engagement.

Once the data is cleaned and structured, I intend to perform sentiment analysis using VADER (Valence Aware Dictionary and sEntiment Reasoner), which will assign positive, negative, neutral, and compound sentiment scores to each comment. These sentiment outputs will be used to explore patterns in how disabled individuals express emotion when discussing their lived experiences online. In the second stage of the project, I will use sentiment scores, specifically the VADER-generated positive, negative, and neutral values, alongside structural features such as upvotes, word count, and character count as the foundation for predictive modeling to determine which factors are most associated with positive or negative expressions of sentiment.

***Sentiment Analysis***

To understand the emotional tone of the Reddit comments, I used VADER to calculate a compound sentiment score for each one. Based on these scores, I labeled each comment as Positive, Negative, or Neutral. This three-way classification gave me a more complete view of how people expressed their experiences, rather than limiting the analysis to a simple positive-versus-negative divide.

Out of the 2,063 comments analyzed, 898 were negative, 886 were positive, and 331 were neutral. The fairly even split between positive and negative sentiment suggests a wide range of emotional expression in how people with disabilities talk about their lives. *Figure 1* shows the overall distribution of sentiment categories.

To explore the most common themes in these discussions, I created a word cloud based on all the comments (*Figure 2*). Words like “disability,” “thing,” “people,” “pain,” “day,” and “work” appeared most frequently, highlighting key concerns and repeated challenges. Many words also reflected emotional or physical struggle, such as “need,” “feel,” “never,” and “support.” The word cloud shows how the conversation often centers on everyday barriers, chronic symptoms, and the desire for understanding.

To better understand the specific language used in emotionally charged comments, I created two separate word clouds—one for comments labeled positive, and one for those labeled negative (*Figures 3 and 4*). These visualizations help highlight the different ways people talk about disability depending on the tone of their message.

In the negative sentiment word cloud, words like “pain,” “can’t,” “time,” “work,” “without,” and “doctor” appeared frequently. These terms reflect recurring themes of physical discomfort, lack of access, exhaustion, and frustration with medical systems or daily expectations. The prominence of words like “life,” “see,” and “stop” suggest the frequent mention of limitations or barriers that interfere with day-to-day activities.

In contrast, the positive sentiment word cloud still includes shared terms like “disability,” “thing,” “people,” and “day,” but also highlights more constructive or supportive language. Words like “help,” “thank,” “understand,” “love,” “feel,” and “hope” suggest moments of connection, gratitude, and resilience. Interestingly, the word “don’t” appears in both clouds—likely reflecting the structure of many comments (e.g., “I don’t mind…” vs. “I don’t get support…”), underscoring that sentiment cannot always be captured by individual words without context.

I also looked at how sentiment was related to upvotes, which reflect how other users engage with and respond to each comment (*Figure 5*). Interestingly, neutral comments received the highest average number of upvotes, followed by negative, with positive comments receiving the fewest upvotes on average. This may suggest that users tend to value comments that are more balanced, nuanced, or thought-provoking, qualities that often fall into the “neutral” sentiment category. It might also reflect how expressions of frustration or realism attract more engagement than overt optimism.

***Predictive Analysis***

After analyzing the emotional tone of the Reddit comments, I built a machine learning model to predict whether a comment had a positive or negative sentiment. Following the assignment guidelines, I filtered out neutral comments to focus only on clearly positive and negative expressions.

I used a Random Forest Classifier for this task. I chose this model because it supports binary classification, works well with both numerical and text-derived features, and reduces overfitting by averaging the results from multiple decision trees. The features I used included the number of upvotes, and the VADER-generated positive and negative sentiment scores for each comment.

I split the data into training and testing sets using a 75/25 split. After training the model, I tested it on the held-out data and evaluated its performance using a confusion matrix and classification report. The model performed very well, achieving an accuracy of 98.6%, with

F1-scores of 0.987 for negative comments and 0.985 for positive comments. Precision and recall were similarly high, indicating that the model made very few classification errors.

As expected, comments with higher positive scores were more likely to be predicted as positive, and those with higher negative scores were more likely to be classified as negative. The number of upvotes had less influence on the outcome, suggesting that community engagement was not a strong predictor of sentiment in this case.

In addition to building a machine learning model using VADER sentiment scores, I also compared VADER to two other popular sentiment analysis tools: TextBlob and Flair. The goal was to evaluate how well these tools aligned with VADER's sentiment labels (positive vs. negative) and to assess their accuracy in sentiment classification.

To do this, I used VADER's output as the baseline and tested how well TextBlob and Flair agreed with it on a binary classification task. For each comparison, I calculated precision, recall, F1-score, and overall accuracy.

The comparison between VADER and TextBlob showed moderate agreement, with an overall accuracy of 72%. TextBlob demonstrated slightly better recall for positive sentiment (0.77), but lower precision for negative sentiment (0.75 vs. 0.69), resulting in balanced F1-scores around 0.71–0.73 for both classes. This suggests that TextBlob can generally detect emotional tone in line with VADER, though it is somewhat more optimistic in its classifications.

The comparison between VADER and Flair revealed more significant disagreement. Flair achieved high recall for negative comments (0.91) but struggled to identify positive ones, with a recall of just 0.32. This produced an overall accuracy of 63%, and a much lower F1-score for positive sentiment (0.46 compared to 0.71 for negative). These results suggest Flair tends to over-classify comments as negative, possibly due to differences in how its underlying model interprets online text.

*Figure 6* visualizes the relationship between TextBlob polarity scores and Flair confidence, colored by VADER sentiment labels. Most positive VADER-labeled comments cluster on the right side of the x-axis (higher polarity), while negative ones lean left. However, the vertical spread in Flair confidence, even among strongly polarized comments, shows inconsistency in how Flair aligns with VADER sentiment categories.

In summary, while both tools offer value, VADER demonstrated the most balanced and reliable performance for this dataset. TextBlob had moderate agreement and may be suitable for broader polarity trends, while Flair's confidence-based outputs appear less aligned with VADER sentiment when applied to Reddit-style text. This comparison reinforces VADER's suitability for short, informal social media content like Reddit comments.

***Conclusions***

This project offered a unique opportunity to explore how people with disabilities describe their everyday experiences through open-ended online discussion. By applying sentiment analysis tools and predictive modeling techniques to Reddit comments, I was able to uncover emotional patterns, recurring themes, and the factors that influence how those sentiments are expressed and interpreted.

One of the most striking findings was the balance between positive and negative sentiment across the dataset, highlighting that conversations about disability are not defined by a single emotional tone. Many users expressed frustration and pain, while others shared moments of resilience, community, and humor. The sentiment distribution shows that disability-related experiences are complex and deeply personal and that the digital space serves as an outlet for those emotions.

I was also surprised by the predictive model's performance. With just a few features, the model was able to classify emotional tone with 98.6% accuracy. This suggests that even simple tools, when paired thoughtfully, can meaningfully interpret emotional expression. It was also interesting to see that neutral comments received the highest average upvotes, hinting at the Reddit community's preference for nuance and balance in responses.

The comparison between VADER, TextBlob, and Flair further emphasized that not all sentiment analysis tools interpret emotional content in the same way. While TextBlob performed reasonably well, Flair's tendency to over-classify comments as negative highlighted the challenges of applying general-purpose models to niche or informal online discussions.
Overall, this project reinforced that language matters and that how people talk about disability reveals not only their emotions, but also the systems, challenges, and relationships that shape their lives.

## *Appendix*

Figure 1

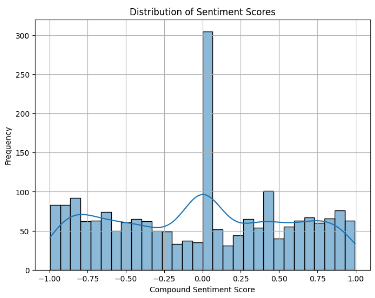


Figure 2

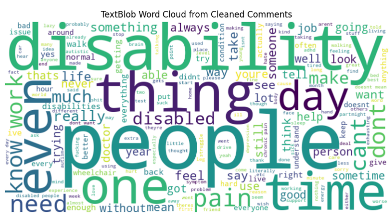


Figure 3


Figure 4

Figure 5

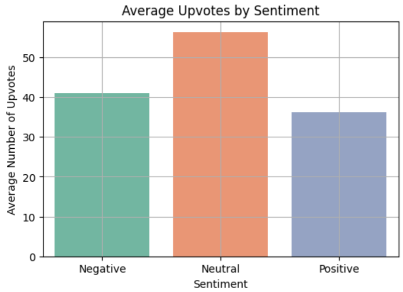


Figure 6