Analyzing Mental Health Discourse on Social Media

Nina Chen
Sarvenaz Hashemiborzabadi

University of Colorado Boulder

Natural Language Processing

Professor Handler

03/05/2024

Introduction

For the group project of the NLP class, we tried to find a dataset that we are interested in its topic, and we can run an analysis on it in order to draw meaningful conclusions. We were looking for a dataset from reliable sources such as PubMed, Google Dataset Search, and Kaggle. Each member of our group came up with a dataset and then we chose one. The dataset that we had a consensus to use was Mental Health Corpus from Kaggle. According to the data description on Kaggle, "The Mental Health Corpus is a collection of texts related to people with anxiety, depression, and other mental health issues." This dataset contains only two columns: one column is the text that was written by someone, and the other column is the label associated with that text, which indicates whether this text is poisonous.

We were interested in analyzing the mentioned dataset because we found it very important to identify the risk factors and the toxic behavior patterns associated with mental health conditions. By having more information about the mental health conditions we can significantly improve the society's mental health. People, especially teenagers, talk about their mental health issues on social media. When social media platforms are able to identify the texts that contain worrying words about the writer's mental health, they can flag harmful interactions between people. Also, tracking these patterns can help mental health professionals quickly identify the potential harmful language. Therefore, it was interesting for us to identify the language that people with mental health concerns such as depression and anxiety use on online platforms.

Research Questions

Before starting our analysis we had many questions about the dataset that we were going to answer based on our analysis. The questions that we had were:

- Is the performance of traditional machine learning models such as logistic regression in detecting depression-related text different compared to deep learning models like LSTM?
- What are the most correlated features with depression in the textual data?
- What is the semantic association between depression-related words and the concepts of 'depression' and 'anxiety' as measured by cosine similarity scores?

When we started analyzing our dataset, we tried various NLP methods and ML algorithms to answer those questions. Some of our attempts were successful and some were unsuccessful. The questions that we could successfully answer based on our codes were about the performance difference between ML and DL methods, the top words that are highly associated with depression, and the semantics association between depression-related words and the concepts of depression and anxiety measured by cosine similarity.

Model Implementation: Unveiling the Impact of Depressive Language Detection Across Domains

Our model designed to identify the depressive language in text could be immensely valuable across various domains. Mental health professionals could utilize it as a screening tool to analyze patient communications, enabling early detection and monitoring of depressive symptoms. Social media platforms could integrate the model to flag posts with concerning language, facilitating timely interventions for users experiencing mental health challenges. Online support communities and helplines could leverage the model to identify individuals in need of immediate assistance. Researchers could analyze large text datasets to gain insights into population-level mental health trends and treatment outcomes. Additionally, individuals could use applications equipped with the model to track changes in their own language patterns, fostering self-awareness and facilitating appropriate support or treatment-seeking behavior.

NLP Methodologies

- **The preprocessing steps** we've implemented to refine our textual data are as follows. Firstly, we've standardized all tokens to lowercase. This helps ensure consistency and reduces the complexity of our dataset by treating 'Word', 'word', and 'WORD' as the same entity. Next, we've removed any URLs present in the text. URLs often carry no meaningful information for our analysis and can clutter our dataset. Then, we've eliminated punctuation marks. While punctuation is essential for grammar, it's usually irrelevant to our analysis. Removing it streamlines our data and ensures punctuation doesn't interfere with our algorithms. Following that, we've removed common stopwords such as 'and', 'the', 'is', etc. These words appear frequently in text but don't contribute much to the overall meaning. By removing them, we focus on the more meaningful words in our dataset. Lastly, we've performed stemming using the Porter Stemmer algorithm. Stemming reduces words to their root form, which helps in consolidating similar words. For example, 'running', 'ran', and 'runs' all reduce to 'run'. This simplification aids in text analysis tasks.

- **Traditional machine learning model - Logistic Regression:**
  Firstly, we divided our dataset into training and testing sets using a common technique called train-test split. This ensures that our model is trained on one portion of the data and evaluated on another to measure its performance. Then, we converted our text data into numerical features using a technique called TF-IDF (Term Frequency-Inverse Document

Frequency) vectorization. This process transforms text data into a format that machine learning algorithms can understand and process effectively. Next, we trained a logistic regression model on the training data. Logistic regression is a type of linear model commonly used for binary classification tasks like ours, where we predict whether a text sample is associated with depression or not. After training the model, we evaluated its performance on the testing data and found that it achieved a certain accuracy score. This score indicates how well the model generalizes to unseen data.

- **The coefficients of the logistic regression model to identify the top 15 features associated with depression :**
  We analyzed the coefficients of the logistic regression model to identify the top features associated with depression. These coefficients represent the importance of each word (or feature) in predicting depression.

- **Cosine similarity:**
  We utilize the top coefficients of the logistic regression model to discern the top 15 features associated with depression. Subsequently, we compute the cosine similarity between each depression-related word and the vectors representing 'depression' and 'anxiety.' Cosine similarity quantifies the cosine of the angle between two vectors, yielding a value between -1 and 1, with higher values indicating stronger similarity. By comparing the cosine similarity scores of each word to 'depression' and 'anxiety,' we can discern the extent of association between these words and the respective mental health concepts.

- **Deep learning:**
  we employed Long Short-Term Memory (LSTM) to tackle a sentiment analysis on textual data. To prepare our data for modeling, we first tokenized the textual data using the Keras Tokenizer. This step involved converting the text into numerical sequences, which the model could understand and process. Next, we padded the sequences to ensure they all had the same length. This step is crucial for deep learning models like LSTM, which require inputs of uniform length.

  With our data prepared, we split it into training and testing sets, allowing us to train the model on one portion and evaluate its performance on unseen data. Our model architecture comprises an embedding layer to convert words into dense vectors, followed by two LSTM layers with dropout regularization to prevent overfitting and a dense layer with a sigmoid activation function for binary classification. We compile the model with the Adam optimizer and binary cross-entropy loss function, suitable for binary classification tasks like ours. To prevent overfitting during training, we employ early stopping, a technique that monitors the validation loss and stops training when it begins to increase, restoring the best weights obtained during training. We then train the model for 20 epochs, monitoring its performance on both the training and validation sets. After training, we evaluate the model on the test set and measure its accuracy.

- **Annotation**:
  We (Sarvenaz and Nina) selected 20 texts from the dataset and annotated whether each text was poisonous. Then, we calculated the Chance Agreement Rate and Observed Agreement Rate.

Results and Insights

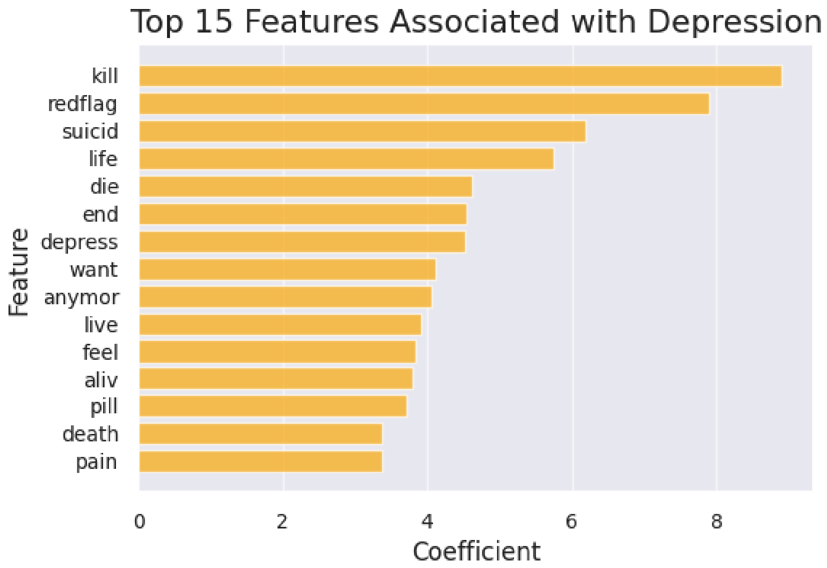
- **Type-Token Ratio (TTR)**

	All Text	Text label 0	Text label 1
Type-Token Ratio (TTR)	0.0360	0.0729	0.0314

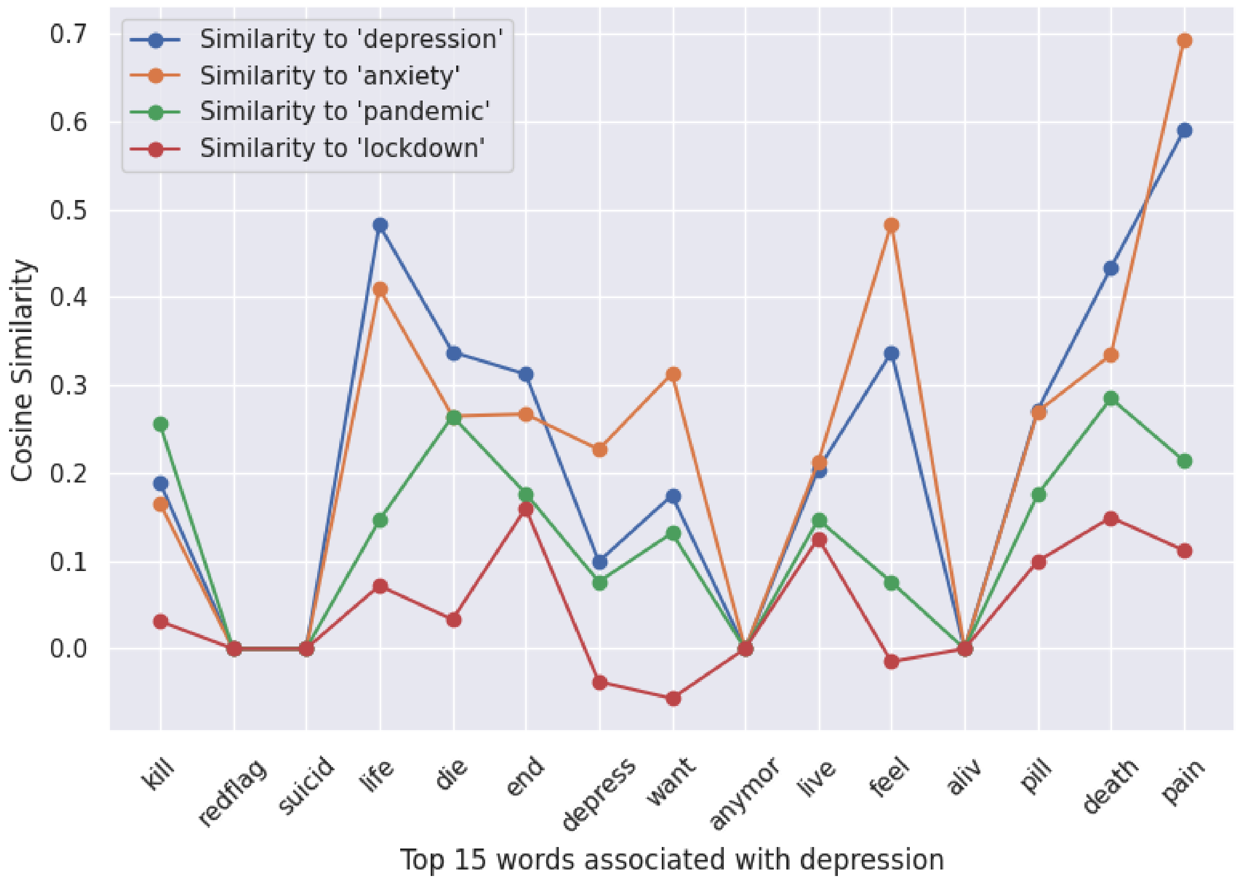
- **Traditional machine learning model: Logistic Regression**
  Training Accuracy: 0.94

  Test Accuracy: 0.93

- **The coefficients of the logistic regression model to identify the top 15 features associated with depression**



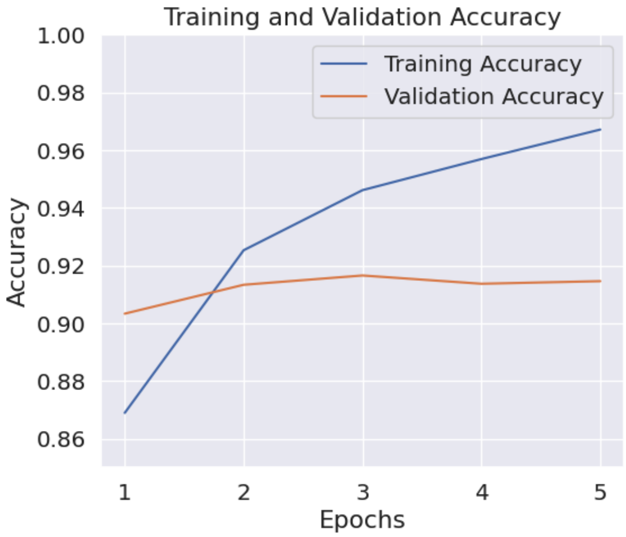
- **Cosine Similarity between Logistic Regression Coefficients and "Depression", "Anxiety", "pandemic", "lockdown" Features**

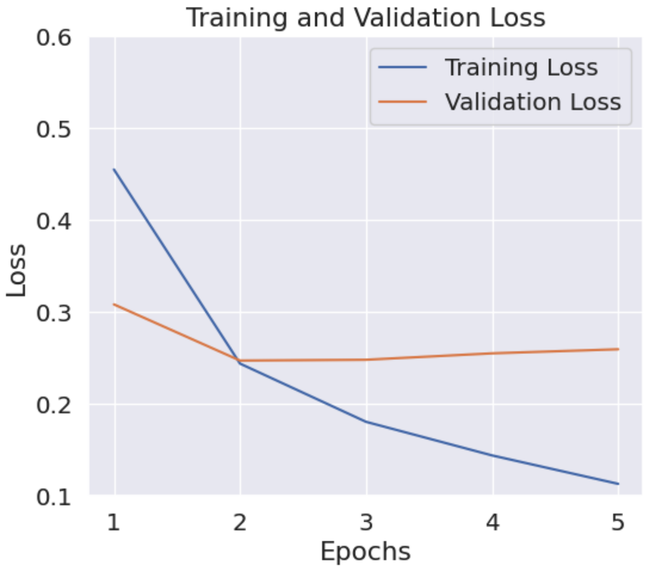


The cosine similarity analysis reveals intriguing insights into the association between specific words and the concepts of 'depression' and 'anxiety.' Notably, words like "life," "die," "end," "feel," "pill," "death," and "pain" exhibit relatively high similarity scores to both 'depression' and 'anxiety,' indicating shared emotional states. However, certain words display distinct associations; for instance, "pain" shows a stronger connection to 'depression' than 'anxiety,' while "feel" appears more closely linked to 'anxiety' than 'depression.' Interestingly, neutral words like "redflag," "suicid," and "anymor" exhibit no significant association with either mental health condition or these words are not present in the glove_word_vectors dictionary.

This information holds significant business value, enabling companies to tailor mental health support services, improve content moderation on online platforms, and refine marketing strategies to better address the emotional needs and concerns of their target audience.

- **Deep learning accuracy: 0.924**





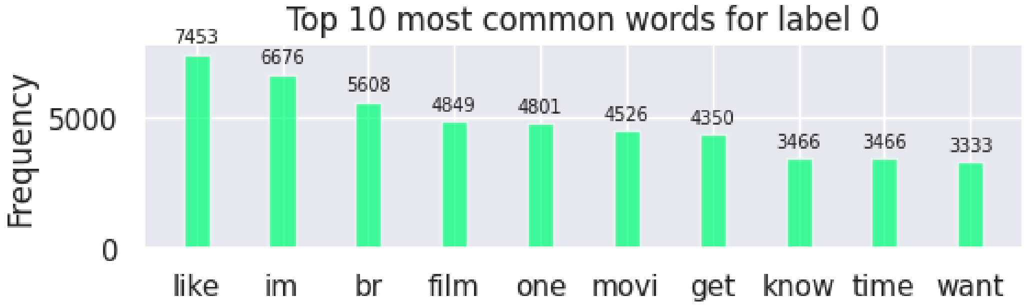
The training and validation results demonstrate a consistent pattern indicative of a well-generalized model. Both accuracy and loss metrics exhibit favorable trends, with training accuracy increasing steadily and training loss consistently decreasing over the epochs. Moreover, the validation accuracy stabilizes around 91.46%, suggesting that the model effectively learns from the training data without overfitting. The validation loss also levels off after initial decreases, indicating the model's ability to maintain performance on unseen validation data. Overall, the close alignment between training and validation metrics suggests that the model successfully generalizes to unseen data, underscoring its robustness and reliability. Continued vigilance and monitoring for overfitting remain prudent practices for ongoing model refinement.

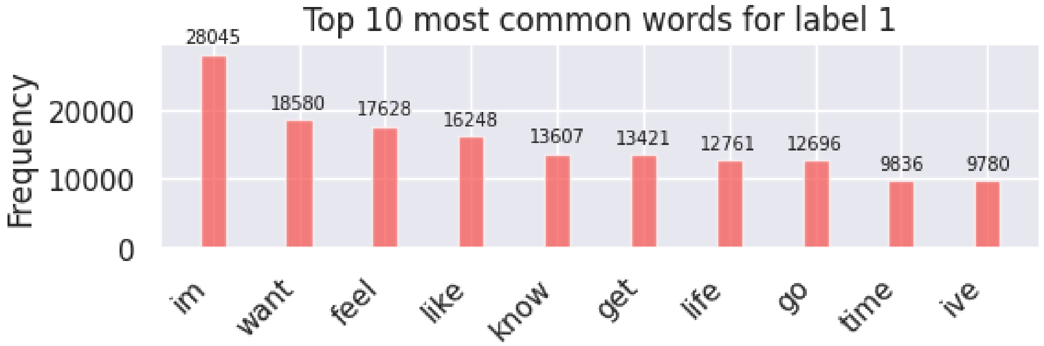
- **Annotation**

  Chance agreement rate: 0.5161

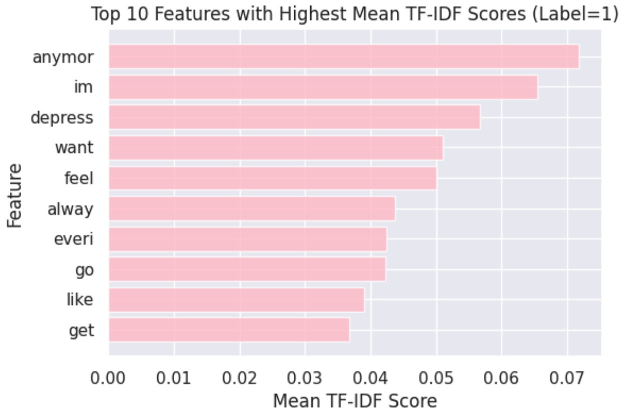
  Observed agreement rate: 0.8947

- **Top 10 most common words for labels 0 and 1**

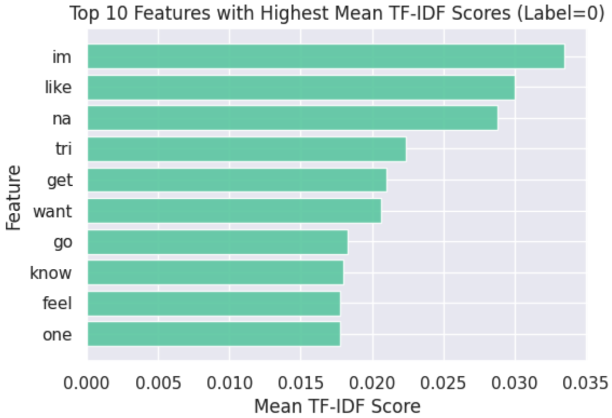




- **TF-IDF**



The top features associated with label 1 "depressive" provide valuable insights into the thematic focus and recurring patterns within the analyzed text. "Anymore" stands out as the highest-ranking feature, suggesting a preoccupation with temporal aspects and potentially indicating expressions of despair or hopelessness linked to ongoing or prolonged experiences. Following closely is the contraction "I'm," highlighting a strong emphasis on personal experiences or perspectives within the depressive context, likely reflecting self-referential thoughts or feelings. Unsurprisingly, the term "depress" appears prominently, underscoring the prevalence of discussions or expressions related to depression. Additionally, words like "want" and "feel" point towards a thematic focus on desires, emotions, and subjective experiences, which are central themes in depressive discourse. The inclusion of words such as "always" and "every" hints at a sense of perpetuity or universality in depressive expressions, indicating persistent or pervasive thoughts, emotions, or experiences. Furthermore, commonly used words like "go," "like," and "get" exhibit elevated TF-IDF scores, suggesting their recurrent usage or significance within depressive discourse. These insights deepen our understanding of depressive language patterns and can inform the development of targeted interventions or support strategies for individuals experiencing depression.



The top features associated with the label "0" (not labeled as depression) provide insights into the prevalent themes and recurring patterns within the analyzed text. "Im" and "like" emerge as the highest-ranking features, indicating a focus on personal experiences and preferences, which are common in everyday language. The term "na" also appears prominently, suggesting informal or colloquial language usage. Additionally, words like "get," "try," and "want" hint at actions, desires, or intentions, which are typical elements of casual conversation. The presence of words such as "go," "know," "feel," and "one" further underscores the varied topics and expressions within the text, reflecting a diverse range of subjects and perspectives. These insights highlight the multifaceted nature of non-depressive language patterns and contribute to a deeper understanding of the textual data.

Conclusion

In this project, we delved into a comprehensive analysis of textual data, focusing on mental health and sentiment analysis. Leveraging traditional machine learning techniques like logistic regression, we identified pivotal features associated with depression and anxiety, complemented by cosine similarity analysis. Additionally, we harnessed deep learning through LSTM for nuanced sentiment analysis. Our findings unveil valuable insights with broad applications, ranging from targeted interventions by mental health organizations to content moderation on social media platforms and sentiment analysis in healthcare. Notably, our model tailored to detect depressive language in text stands as a potent screening tool for mental health professionals, facilitating early detection and monitoring of depressive symptoms. Its integration into social media platforms could streamline flagging of posts with concerning language, enabling timely interventions for users grappling with mental health challenges. Furthermore, online support communities and helplines could leverage this model to swiftly identify individuals in urgent need of assistance, amplifying the reach and effectiveness of their services.

References

https://www.kaggle.com/datasets/reihanenamdari/mental-health-corpus

Report Writing Assistance: ChatGPT is utilized to help with the elegance of speech with the original idea remaining the team’s own.