"Mental Health Analysis"

CourseCode:CS-538A

Team Members:
Akram Pathan (apathan3@stevens.edu)
S. Aditya Reddy (sreddy5@stevens.edu)

Abstract:

In this research endeavour, the central aim is to transcend simplistic binary classifications prevalent in the analysis of online discourse, particularly on platforms like Reddit. Instead of merely distinguishing between casual conversations and discussions related to mental health, the project aspires to undertake a comprehensive and nuanced examination of mental health ailments and their varying degrees of severity. By leveraging textual data extracted from Reddit, the study endeavours to go beyond mere identification of the type of mental health condition discussed in posts and delve into a more intricate analysis of the severity spectrum associated with these conditions. The overarching objective is to develop a sophisticated system that not only categorizes posts based on the nature of mental health issues but also incorporates a nuanced assessment of their severity levels. Furthermore, the project seeks to create a robust framework capable of assigning distinct labels, going beyond generic categorizations, and even extending to the identification of potential risk factors, such as assessing the risk of suicide within the discussed mental health context.

In order to achieve this multifaceted analysis, the research will employ advanced natural language processing techniques and machine learning algorithms to extract meaningful patterns from the vast and diverse textual data available on Reddit. By discerning subtle nuances in language and content, the project aims to contribute to a more refined understanding of mental health discussions in online communities. This initiative holds significant promise not only in advancing our comprehension of mental health-related conversations but also in the development of proactive systems that can identify and address potential risks, thereby enhancing the overall well-being of individuals participating in online mental health discourse.

GitHub Link: https://github.com/akrampathan07/Deep-Learning-Project

Introduction:

The critical role of mental health in an individual's overall well-being is indisputable; however, today, it often does not receive the attention and importance it deserves. Mental health disorders persist as significantly underreported phenomena, contributing to a pervasive lack of awareness and understanding. Global statistics reveal that an alarming 1 in 10 individuals grapple with mental health disorders, underscoring the urgency to address this pervasive issue. Unfortunately, the societal stigma surrounding mental health prompts many individuals to remain reticent about their struggles, even with their closest peers.

Against this backdrop, social media emerges as a unique and powerful platform where individuals feel compelled to express their thoughts and experiences more openly. Recognizing the potential of leveraging this vast repository of user-generated content, our research employs advanced Natural Language Processing (NLP) techniques, including Long Short-Term Memory (LSTM) and the Bert Algorithm, to glean insights into people's mental states based on their writings on platforms like Reddit. The overarching goal is to harness these insights to establish online pathways connecting users to relevant health information and support resources, fostering individualized interventions.

Despite the wealth of information available on social media, there exists a considerable gap in research pertaining to the nuanced understanding and classification of distinct mental health diseases and their severity. This project seeks to bridge that gap by moving beyond traditional binary classifications, such as distinguishing casual conversations from those related to mental health. Instead, our approach involves a meticulous analysis of textual data extracted from Reddit, aiming not only to identify the specific type of mental health ailment discussed in posts but also to assess the severity spectrum associated with each condition. This endeavour culminates in the creation of a sophisticated system capable of assigning distinct labels, with a specific focus on gauging the risk of suicide within the context of the discussed mental health issues.

By advancing beyond conventional analyses, this research endeavours to contribute significantly to our understanding of mental health conversations in the digital realm. Moreover, the outcomes of this project hold the potential to inform the development of proactive and personalized interventions, thereby fostering a more supportive online environment for individuals navigating the complexities of mental health.

Contribution:

Aditya Reddy played a pivotal role in our project by overseeing crucial tasks related to data handling and preparation. His responsibilities included data extraction, a foundational step in our research, ensuring the acquisition of relevant information from the chosen platform, Reddit. Aditya also spearheaded the pre-processing phase, where raw data was refined and organized to make it suitable for analysis. Additionally, he took charge of training the Long Short-Term Memory (LSTM) model, a key component of our research, and contributed significantly to the development of both the project report and presentation slides.

On the other hand, Akram Pathan assumed a key role in data modeling, contributing his expertise to the structuring and organization of our dataset. His responsibilities extended to assessing the need for any further adjustments or refinements in the data, ensuring its optimal suitability for our research objectives. Akram also led the training of the BERT model, a critical aspect of our advanced Natural Language Processing techniques. Moreover, he made substantial contributions to both the project report and presentation slides, consolidating the outcomes and insights derived from the BERT model.

In summary, Aditya Reddy's focus on data extraction, pre-processing, and LSTM model training complemented Akram Pathan's expertise in data modeling, refining data, and BERT model training. Together, their collaboration significantly enhanced the robustness of our project, covering essential aspects from data preparation to advanced modeling techniques, and ultimately contributing to the comprehensive project report and presentation slides.

Data:

Our dataset serves as the foundational cornerstone of our research, meticulously curated through the utilization of Reddit's PRAW API. This comprehensive dataset comprises diverse posts sourced from specific subreddits, each meticulously selected to represent distinct facets of mental health discourse. The chosen subreddits include those dedicated to specific mental health conditions such as depression, anxiety, Bi-Polar Disorder (BPD), and SuicideWatch. Additionally, we incorporated the r/mentalhealth subreddit, which encapsulates general discussions on mental health, providing a broader context. The inclusion of r/CasualConversation, although not aligned with mental health discussions, is deliberate. This subreddit introduces a variety of data to our models, serving as a crucial element for ensuring the robustness and generalization capabilities of our analyses.

The dataset encapsulates approximately 12,600 posts, each labeled with the corresponding subreddit to facilitate supervised learning. These labels categorize posts into specific mental health conditions or general discussions, offering a rich and diverse array of textual data for our analysis. The dataset's columns of interest include the post's Title, providing a succinct summary of the content; the Body, which contains the detailed text of the Reddit post; Score, representing the net engagement with the post calculated as the number of upvotes minus the number of downvotes; Upvote Ratio, offering insights into the post's popularity by indicating the ratio of upvotes to the total number of votes; and Subreddit, the categorical label identifying the specific subreddit to which the post belongs.

These columns collectively form the crux of our dataset, enabling us to conduct a nuanced analysis of mental health discussions on Reddit. The richness of information captured, coupled with the deliberate inclusion of diverse subreddits, empowers our models to glean insights into the multifaceted nature of mental health discourse, ultimately contributing to the depth and comprehensiveness of our research findings.

Some of the columns of interest are:
- Title: Title of the post
- Body: The content of the Reddit post
- Score: It is the number of upvotes minus the number of downvotes
- Upvote Ratio: It is the number of upvotes divided by total number of votes
- Subreddit: Contains labels depicting which subreddit the corresponding post belongs to.

Methods:

The methodology employed in this research encompasses a comprehensive set of steps aimed at refining and analyzing the textual data extracted from Reddit, specifically focusing on mental health-related discussions. The following detailed procedures were undertaken in the process:

1. **Data Cleaning:**

- To ensure the integrity of our dataset, missing values were addressed by removing irrelevant entries not directly related to mental health conditions. A Reddit-specific preprocessing step was conducted to eliminate markdown formatting from the text retrieved via the PRAW API, ensuring a standardized and clean dataset for subsequent analysis.

2. **Language Identification and Filtering:**

   Language identification was performed using the langdetect package, with a focus on retaining only posts in English. Stop words and punctuation marks were then systematically removed from the textual data, streamlining it for further analysis.

3. **Textual Data Processing:**

   Stemming and lemmatization techniques were applied to the textual data, with a discernible improvement observed with lemmatization, enhancing the accuracy and relevance of subsequent analyses.

4. **Vectorization Techniques:**

   The textual data underwent multiple vectorization techniques, including the utilization of CountVectorizer to convert it into a matrix of token counts, unveiling prevalent unigrams, bigrams, and trigrams. TF-IDF Vectorizer was employed to represent the importance of terms in the documents, and Word2Vec was used to create word embeddings, capturing contextual information within full sentences.

5. **Sentiment Analysis:**

   A crucial aspect of the methodology involved addressing the challenging task of classifying the severity of mental health conditions through sentiment analysis. Sentence polarity, indicating the sentiment of individual sentences, was calculated by combining the polarities of constituent words, providing a nuanced understanding of the posts.

6. **Visualization and Analysis:**

   The research delved into visualizing and analyzing polarity distributions for various mental health states, offering insights into sentiment trends within the dataset.

7. **Clustering:**

   Clustering was explored using Principal Component Analysis (PCA) to reduce dimensionality. Various classification models, including Naive Bayes, Logistic Regression, and RandomForest Classifier, were implemented with different vectorization techniques to identify patterns and relationships within the data.

Issues and Insights:

The methodology acknowledged and addressed issues with clustering, particularly in posts related to different mental health conditions. Limitations of Bag of Words (BoW) in capturing contextual information were recognized, emphasizing the challenge of common words across different mental health categories.

Topic Modeling:

To overcome challenges related to the commonality of words and the need for context, topic modeling with Latent Dirichlet Allocation (LDA) was adopted. This approach facilitated the modeling of posts based on the severity of mental illness with fresh labels, revealing key keywords within each cluster.

In essence, the methodology employed a multi-faceted approach, combining data cleaning, linguistic analysis, vectorization techniques, sentiment analysis, clustering, and topic modeling to unravel intricate patterns and insights within the rich textual data from Reddit.

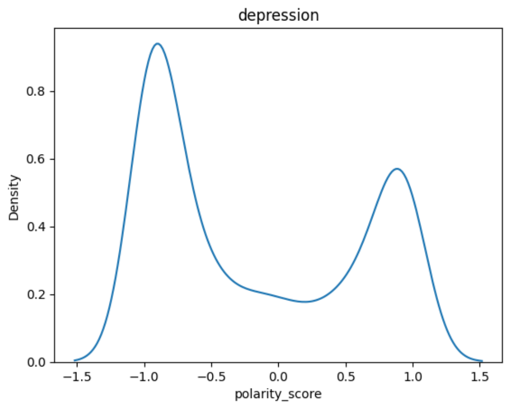


Fig. (a)

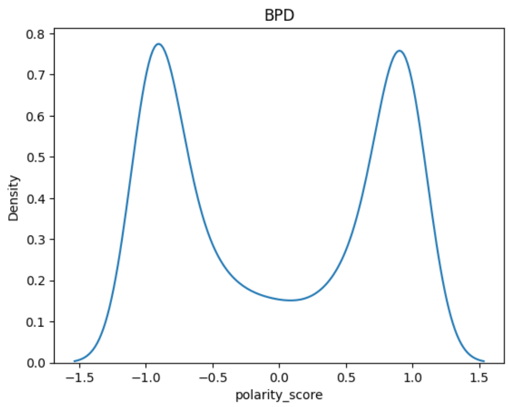


Fig. (b)

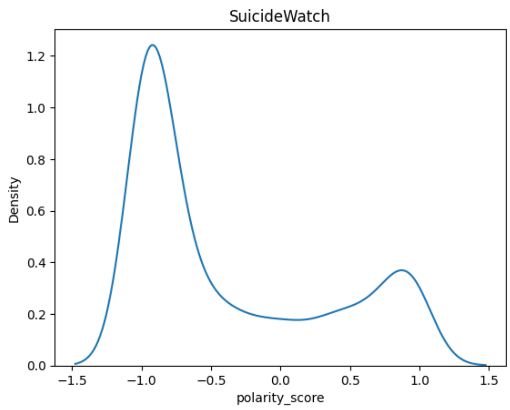


Fig. (c)

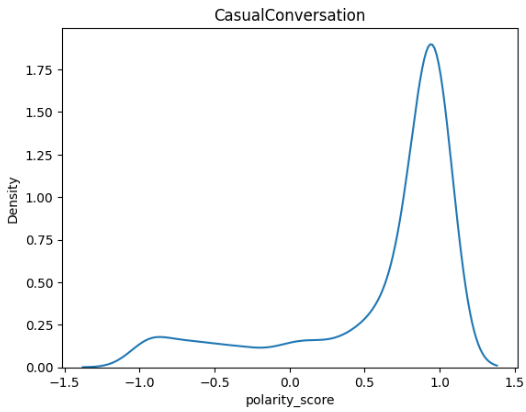


Fig. (d)

From the Figure (a), We can see that the polarity of depression is highly peaked towards negative (-1) which represents the negative sentiment and has a peak towards the positive (+1) which is not as high as negative sentiment. The small peak towards positive can be referred to as, them (people in depression) talking about situations they were happy about in the past.

From the Figure (b) we can see that the polarity of BPD (bipolar disorder) is bi-modal. The peaks in the graph are at the positive (+1) and negative (-1) sentiment (Extremes). This represents that the sentiment of people who are suffering from bipolar disorder have a possibility of being at positive sentiment and at the negative sentiment (which aptly classifies them as Bipolar).

From the Figure (c), we can see that the polarity of SuicideWatch is highly negative (-1) which means the sentiment is highly negative, and very less towards positive (+1).

From the above Figure (d), we can see that the polarity of Casual Conversation is more tending towards positive (+1) which means it has more positive sentiment and very less towards negative (-1).

Following topic modeling, we require a model that classifies the comment based on context and captures the meaning. For this categorization assignment, we employed the Long Short-Term Memory (LSTM) algorithm.
We chose the LSTM model since we had to deal with data that needed to be ordered. LSTM models are known to outperform Recurrent Neural Networks (RNN) models due to the problem of exploding and disappearing gradients in RNN models. A'memory cell' in LSTM units may keep information in memory for long periods of time. A collection of gates controls when information enters the memory, when it is output, and when it is erased. They can learn longer-term dependencies thanks to this architecture. In simple words, it performs well in capturing the context in the sentence.

```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 embedding (Embedding)       (None, 2553, 128)         2546944

 batch_normalization (Batch  (None, 2553, 128)         512
 Normalization)

 activation (Activation)     (None, 2553, 128)         0

 dropout (Dropout)           (None, 2553, 128)         0

 lstm (LSTM)                 (None, 2553, 256)         394240

 dropout_1 (Dropout)         (None, 2553, 256)         0

 dense (Dense)               (None, 2553, 4)           1028

=================================================================
Total params: 2942724 (11.23 MB)
Trainable params: 2942468 (11.22 MB)
Non-trainable params: 256 (1.00 KB)
```

Embedding Layer: The vectorized representation of the phrases we give as input to the LSTM model makes up the embedding layer. Padding can be used to feed the model input sequences of the same length.
Batch Normalization: Batch Normalization layer is used to normalize the inputs. Batch normalization applies a transformation that maintains the mean output close to 0 and the output standard deviation close to 1.
Activation Layer: An activation function in a deep learning model defines how the weighted sum of the input is transformed into an output from a node or nodes in a layer of the network. The choice of activation function in the hidden layer will control how well the network model learns the training dataset. The choice of activation function in the output layer will define the type of predictions the model can make.
Dropout: Dropout is a regularization method where the input and recurrent connection to the LSTM units is excluded from the activation and weight updates probabilistically while training a network. This helps in reducing overfitting and improving the overall model performance.
Dense Layer: A fully connected layer (all nodes of the current layer are connected to all other nodes in the next layer). It often follows LSTM layers and is used for outputting a prediction.

Tools and Technologies:

- Google Colab
- Reddit (for Data Extraction)

Results:

LSTM (before Word Embedding)

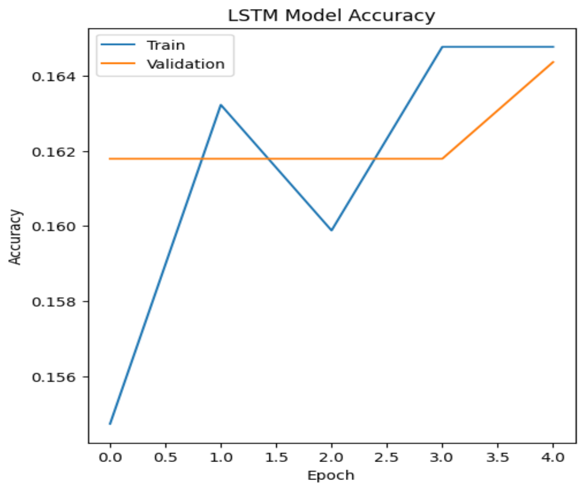


Fig.(e)

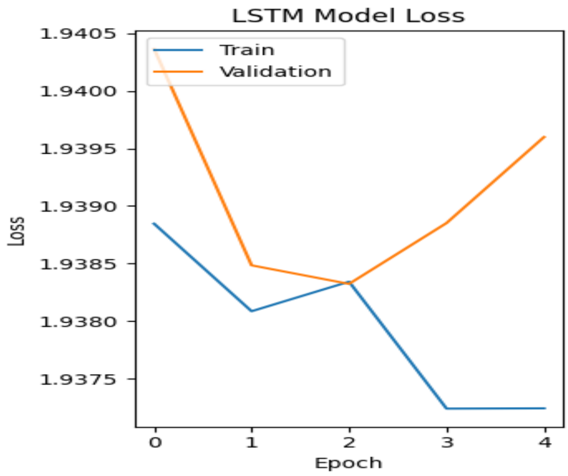


Fig.(f)

The results of our LSTM model are shown in Figures (e) and (f). We trained the model using 5 epochs and attained a test accuracy score of about 16.4%.

BERT (before Word Embedding)

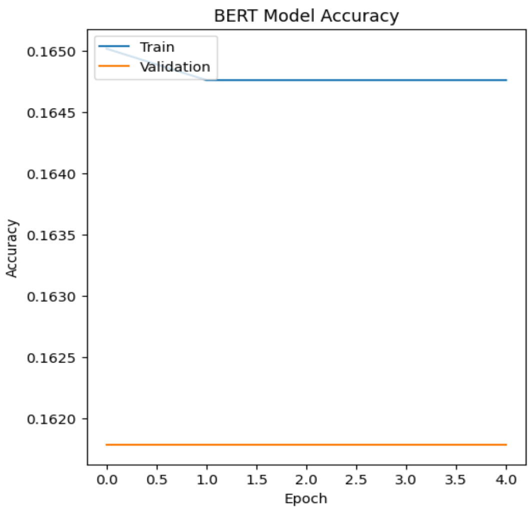


Fig.(g)

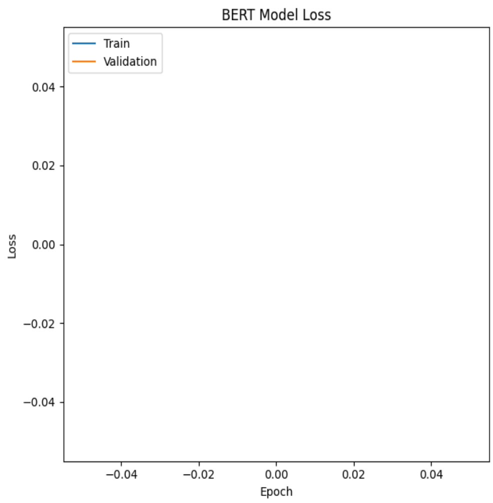
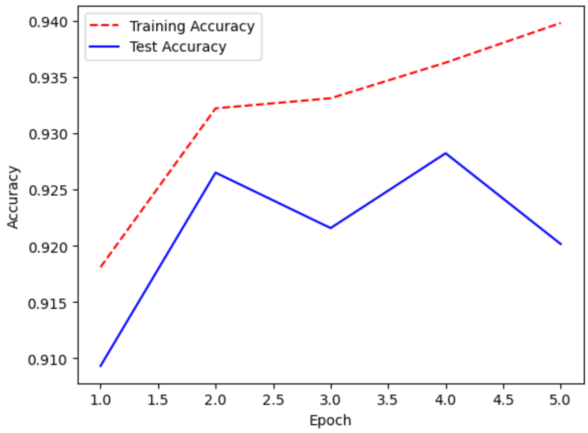


Fig.(h)

The results of our BERT model are shown in Figures (g) and (h). We trained the model using 5 epochs and attained a test accuracy score of about 15%.

LSTM (After Word Embedding and LDA)



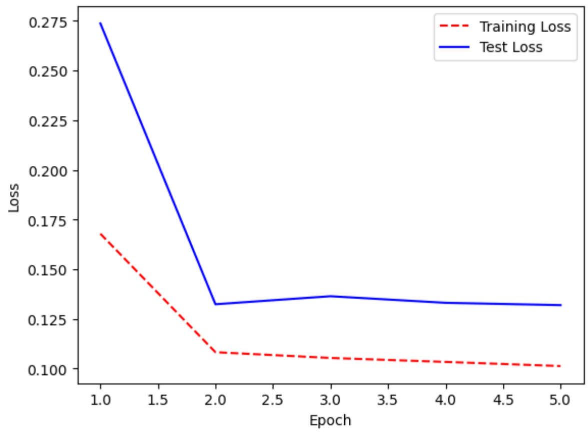


Fig. (i)

Fig. (j)

The results of our LSTM model are shown in Figures (i) and (j). We trained the model using 5 epochs and attained a test accuracy score of about 92%.

BERT (After Word Embedding and LDA)

Using BERT for longer texts, especially those with a length of 2553 tokens, can be challenging. BERT's maximum token limit. The original BERT model has a maximum token limit of 512, and this includes both input and output tokens. To handle longer texts, you need to preprocess and truncate or split your text into segments that fit within this limit.

Here are some common approaches to deal with longer texts when using BERT:

1. Truncation:
    - Simply truncate your text to the maximum token limit. This means you might lose information from the end of your text.
    - Keep the most relevant part of your text and remove the rest
2. Sliding Window:
    - Create overlapping segments of your text using a sliding window approach. This involves creating multiple segments of fixed length and allowing them to overlap, ensuring that no information is lost.
    - Predictions from different segments can be combined (e.g., averaged) for results.
3. Chucking:
    - Divide your long text into smaller chunks and process each chunk separately.
    - Be cautious about the context at the boundary between chunks.

We were facing errors in implementing the BERT model as the data consists of text with a MAX length of 2549 whereas the BERT model could only tokenize up to 512 tokens. We have tried implementing it with Truncating, slicing window and chucking which has caused a lot of data loss and missing major information used for classification of data.

Future Work:

In the future, we plan to implement Longformer which is a transformer-based model designed to handle longer sequences of text efficiently. It extends the self-attention mechanism used in transformers to work on global attention patterns, allowing it to handle longer sequences without a quadratic increase in complexity. Also, Transformers are known to perform better in capturing the context in a sentence. We also plan to deploy this Machine Learning model on Heroku, moving a step forward in making it closer to implementing this on social media platforms.

Conclusion:

The major motivation for this study was to examine people's mental health based on their social media posts. Mental health is extremely important and is undervalued today. Rather of approaching this as a binary classification problem, we went a step further and rated the severity by assigning labels depending on the post's content. We used the LSTM model for this categorization, although we originally ran into overfitting issues.

To avoid training the model for any longer than necessary, we sought to make the network shallower and limit the number of epochs. On the test data, we reached a 92% accuracy.

References:

- Latent Dirichlet Allocation, https://radimrehurek.com/gensim/models/ldamodel.html
- Long Short-Term Memory network, https://colah.github.io/posts/2015-08-Understanding-LSTMs/
- Principal Component Analysis, https://scikitlearn.
- org/stable/modules/generated/sklearn.decomposition.PCA.html
- Batch Normalization, https://machinelearningmastery.com/batch-normalization-for-training-ofdeep-neural-networks/
- Bag of Words, A Simple Explanation of the Bag-of-Words Model - Towards
- ...https://towardsdatascience.com › a-simple-explanation-...
- Depression Reddit, https://www.reddit.com/r/depression/
- Anxiety Reddit, https://www.reddit.com/r/Anxiety/
- Bi-Polar Disorder Reddit, https://www.reddit.com/r/BPD/
- Mental Health Reddit, https://www.reddit.com/r/mentalhealth
- Suicide Watch Reddit, https://www.reddit.com/r/SuicideWatch/
- Casual Conversation Reddit, https://www.reddit.com/r/CasualConversation/