CSCE 5290 - NATURAL LANGUAGE PROCESSING

SUICIDE IDEATION DETECTION USING NLP ANALYSIS

Group: 15

Team Members:

1. Rohit Kolli (11561245)

2. Mittalben Jay Bhatiya (11514905)

3. Vandana Reddy Neelapureddygari (11550895)

4. Ananya Samala (11527196)

Motivation:

Social media plays a significant role in the 20th century, as technology is increasing people prefer to share mostly about their personal lifestyle on media platforms. We have many sites like twitter, reddit and many others. They mostly tweets, blogs to communicate suicidal tendencies.

We observe these clinical suicidal symptoms and online behavior on social media. We detect these suicidal ideas posted by the users using machine learning techniques and Natural Language Processing (NLP) methodology with the help of some set of feature groups like n-grams, bag-of-words, Linguistic Inquiry Word Count (LIWC) for diagnosing depression in user.

Significance:

The project "Suicide ideation detection" is important for several reasons:

To begin with, suicide is a major public health concern, with millions of people worldwide attempting suicide every year.

Suicide ideation, or thoughts of suicide, is a key risk factor for suicide, and detecting these ideations early can help prevent suicide attempts.

It is imperative to accomplish the purpose of this project for several reasons:

1. Identifying individuals with suicidal ideation early can help prevent suicide attempts.
2. NLP analysis can help detect suicide ideation early by analyzing text data from various sources including social media posts, online forums, and chat messages. NLP techniques can improve the accuracy of suicide ideation detection by analyzing the words used and the context and sentiment of the language used.
3. Suicide ideation detection through NLP analysis is a non-intrusive and cost-effective method compared to traditional methods of screening, such as face-to-face interviews or psychological assessments and help identify individuals who may be at risk of suicide and provide targeted intervention and support.

Early detection of suicide ideation using NLP analysis can help identify individuals who may be at risk of suicide and provide timely support and intervention, preventing suicide attempts and saving lives. Overall, NLP analysis can improve the accuracy and efficiency of suicide ideation detection, making it a valuable tool for suicide prevention efforts.

Objectives:

Building NLP Model: The objective is to build a model that is accurate to the pattern in which we use and predict the likelihood of suicide.

Evaluating Model: This objective is to examine the accuracy and effectiveness of the NLP Model. This involves in testing model on separate samples to determine its ability to correct the identification of patterns for Suicides.

Dataset Description:

Here, we will take the data from twitter and reddit in which we will be having 3000 Tweets and reddit data was collected by querying the keywords like, depression, anxiety, end of my life etc.

Tools and libraries:

Here in this Project, we will be using some natural language processing methods like Text classification, Text Summarization, Sentiment Analysis, POS tagging etc.

The libraries which we will be installing are Spacy, TextBlob, NLTK, and packages Punkt, Stopwords, Wordnet, etc.

Features:

Focusing on the current world examples related to death ratio because of the suicide and depression. Right now, in the US (United States) the most affected disease for people is depression. Our main goal is to identify the reasons and try to provide the platform for detecting the mental health condition of the people.

For achieving our goal, we are using some techniques as follows:

Text Classification:

Text classification is also known as text tagging and text categorization. It is used for processed text based on categories. However, here we are using text classification to add text in to predefined categories.

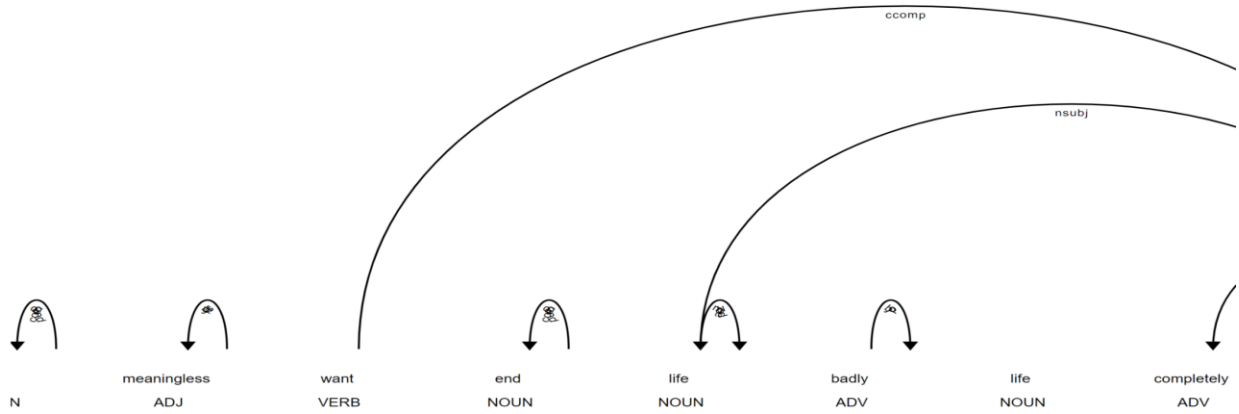
Text Summarization:

Text summarization is the process of paraphrasing the whole lengthy paragraph into some sentences or some meaningful text. In this approach, we will use data from social media and summarize the text into normal form for better and efficient data.

Sentiment Analysis:

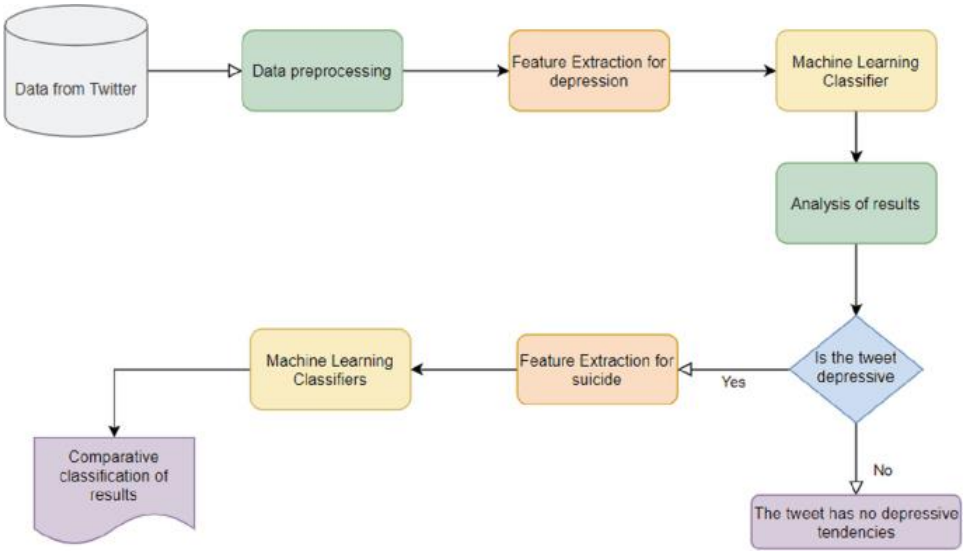
Sentiment analysis is the process of finding the text based on emotions. There are three types of sentiment analysis Positive, Negative, Neutral. Here we are getting posts from twitter and preprocessed the data and now checking the emotions behind the text uploaded on social media. However, sometimes it is not possible to find the sentiment based on text but here we are prediction based on other collected words. Focusing on the above image, all the words are based on the emotions posted on social media so, in conclusion, we are identifying the words and training our model for that.

POS Tagging:



POS tagging is the process of tag the words based on the corresponding part of speech such as verbs, noun, adjectives etc. Here, we refer to pos tagging to get the clearer meaning of the part of speech and text posted on social media.

Visualization:



References:

www.sciencedirect.com/science/article/pii/S2667096822000465

https://monkeylearn.com/sentiment-analysis/#:~:text=Sentiment%20analysis%20(or%20opinion%20mining,feedback%2C%20and%20understand%20customer%20needs.