DSC680 - Depression or cyberbullying sentiments from twitter

Joshua Burden

Bellevue University

03/19/2023

Abstract

In the first term project I chose to examine the topic of depression and cyberbullying using various research papers and datasets found from Kaggle and other resources. The purpose would be to extract meaningful insights into the data by exploring the datasets, and to build a meaningful model to find NLP sentiments that can help predict cyberbullying and early warning indications of depression through social media.

Business Problem

Depression is one of the most prevalent mental illnesses that can be very debilitating, disabling, defeating, but is very curable. It is underwhelmingly undiagnosed, and has a tendency to be created or otherwise increased in intensity due to either online harassment, or negative feedback loops perpetuated from various forms of social media such as facebook, instagram, or twitter. With the current downsizing and purchases of platforms, there is a combination of conversation being had if social media is free, and safe. I am looking at the data points of social media's safety as there is plenty of evidence to suggest that social media helps increase the amount of depression that could be encouraged due to the amount of cyber bullying.

Datasets

There are several datasets that I have found and I think they contain enough data to be able to form an educated opinion based on facts of whether social media is a breeding ground of perpetual cyberbullying resulting in an increased bout of depression. One dataset I found contains a set of over 1,600,000 tweets extracted using the twitter API. There seems to be

potential means of creating a classification system of positive, neutral, or negative comments from the tweets.

Another set of data shows another set of tweets that could be used to create an NLP classification system of cyberbullying tweets using 46,017 tweets. These two datasets could be used to build quite a bit of different models such as a text sentiment analysis engine, and exploratory analysis using word clouds, predictive text models, Long Short term memory networks, and other such Recurrent neural networks.

Summary of Methods

As this is still exploratory this section will need to be created once the data has been cleaned, parsed, explored, and the methods will need to go here later. There is a high likelihood since there will be sentiment analysis, Keyword extraction, Topic Modeling, and other methods that relate to NLP. With the data that is extracted there is also a high likelihood of using graphs such as histograms, barcharts, and word clouds to drive a more visual aspect of the data.

Ethical Considerations

Depression and cyberbullying examination through datasets using api's such as twitter/reddit/facebook need to consider the following metrics and ethical questions:

1. Respect for laws and rules
2. Justice and integrity of the individuals or groups being targeted and those who are persecuting the victims
3. Honesty and altruism in the dataset manipulations and storytelling
4. Respecting the mental and bodily autonomy and confidentiality of those using the platforms

With this in mind, the data should be anonymized in such a way that would respect the individual's privacy. Though the 1st amendment allows for the internet spaces such as facebook, twitter, reddit and other platforms to be a public forum, it is imperative that these tweets and posts are anonymised in such a fashion as to protect any victims from further points of contention while also respecting an individual's right to free speech and self expression. There is no plan for diagnosis either, so performing or using these datasets as means to help either prevent causing more harm, or perpetuating more bullying or depression.

Challenges/Issues

Some of the data sets seem to be older than 2020. There might be an issue getting a more current data set, as API's that were available in the past might not be available for the public now. There is also a challenge of sorting through the data and finding meaningful insights into the data that produce the kinds of data this project is looking to produce. There could also be plenty of opportunities that produce a sentiment that seems like either a cyberbullying or depressive message, but could in fact not be an accurate assessment. A training model with wrong metrics that it is testing against could produce results that are not what the story is trying to tell.

References

https://www.kaggle.com/datasets/andrewmvd/cyberbullying-classification

https://www.kaggle.com/datasets/gargmanas/sentimental-analysis-for-tweets

https://www.kaggle.com/datasets/kazanova/sentiment140

https://medium.com/swlh/detecting-depression-in-social-media-via-twitter-usage-2d8f3df9b313

https://www.who.int/news-room/fact-sheets/detail/depression

https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0012948

https://www.semanticscholar.org/paper/Depressive-Moods-of-Users-Portrayed-in-Twitter-Park-Cha/8dd58913bd343f4ef23b8437b24e152d3270cdaf?p2df

https://www.pnas.org/doi/10.1073/pnas.1802331115

https://arxiv.org/pdf/1804.07000.pdf

http://www.munmund.net/pubs/icwsm_13.pdf

https://www.nature.com/articles/s41598-017-12961-9

https://www.pewresearch.org/internet/fact-sheet/social-media/