Sentiment Analysis of Tweets to Classify Depressive and Negative Feelings in Dutch Twitter Users During the 2020 Global Pandemic

Kira Breeden

June 5, 2021

kmbreeden21@gmail.com

Abstract

Depressive disorders are becoming increasingly prevalent in today's society and with the emergence of the 2020 COVID-19 pandemic, we have a unique opportunity to investigate how global traumatic experiences may affect mental health. The extreme popularity of social media in our modern world provides us with an intricate look into the minds and opinions of the planet. In particular, using artificial intelligence methods such as natural language processing and sentiment analysis, we can observe the general emotions being shown by social media users and we can link these feelings to possible depressive tendencies. This paper investigates indications of depressive tendencies found in Dutch tweets from 2020. We evaluate the performance of different sentiment analyses classification algorithms. We specifically used the existence of stressful life events to inform the categorization of sentiment that may lead to depression. Experimental results show that we are able to achieve a 70% accuracy when performing sentiment analysis on these tweets. These results indicate that the algorithms investigated in this study may be used to identify users displaying particularly negative sentiment that may indicate depressive tendencies.

Keywords: Sentiment Analysis, Recurrent Neural Networks, Long Short-Term Memory, Topic Analysis, COVID-19, Depressive Disorder.

Introduction

Throughout 2020 the world experienced a massive global pandemic that caused worldwide economic, lifestyle, and mental disruption. This COVID-19 pandemic has been extremely grueling mentally and physically for many, but it has also provided the data world with a fascinating environment for data analysis: a worldwide traumatic event. This unprecedented event has given us a new and strange look into how people across the world react to global traumatic events.

With these new global events, it is increasingly important to harness powerful analytics methods to help understand human emotional responses and the mental health effects of such massive disruptions to society. Fortunately, we have had a massive increase in the use of social media since its invent in the late 1990s (Maryville Online 2021). There are over 675 million users on LinkedIn alone as of 2020 (Maryville Online 2021). Social media allows people to express themselves to the public, to their friends, to colleagues, etc through the internet, providing a special look into the minds and opinions of the world. During a global pandemic where physical interaction between people was highly discouraged, connecting online became even more important. Many users of multiple different social media platforms have stated that their use of these outlets has increased since the beginning of the COVID-19 pandemic (Digital Commerce 360 2021).

Another convenient aspect of the growth of social media is the resulting massive amounts of analyzable data. In particular, platforms such as Twitter have built-in APIs that help researchers in the data mining of tweets across the world. This large amount of social media data provides us with a relatively all-encompassing look at the mindset of the world. In a situation such as the COVID-19 pandemic, we were faced with a traumatic event that most of the world's generations had never experienced. In other words, the social media data from this time is completely unique and shows a snapshot in time of a worldwide trauma.

Artificial Intelligence (AI) has been used increasingly frequently in order to analyze this mass amount of social media data at our fingertips and specifically, Sentiment Analysis can be monumentally useful in the mental health sphere (Lin 2020). AI can be used for diagnoses, identification of symptoms, treatment, etc in psychiatry and clinical psychology (Lin 2020). The use of natural language processing (NLP) in social media post sentiment analysis can provide us with a close look at what our public thoughts tell us about ourselves and the world.

In this study we intend to perform sentiment analysis on tweets from 2020 in order to categorize tweets based on negative or depressive sentiment. The classification labels of these tweets will be based on existence of words that may indicate stressful life events — like illness, trauma, abuse, depression, etc (Harvard Health Publishing 2019).

Literature Review

Natural Language Processing (NLP) is a common way of performing sentiment analysis on textual data. In these cases, text data is analyzed according to certain terms and topics that indicate certain connotations of sentiment. This method is used often in today's society for problems such as product review analysis, movie review analysis, customer email satisfaction, etc (Lane et al. 2019). There are many textbooks and references in modern society that provide strong examples of how the basics of sentiment analysis work for items such as movie reviews in the realm of NLP (Lane et al. 2019). In these methods, it is most common to use word markers of emotion to indicate and determine the sentiment of the text excerpt.

The case that will be discussed and investigated in this paper, is the use of sentiment analysis to classify sentiment into different categories. These categories are going to be informed by context clues and words found often in situations that may indicate depressive tendencies (Harvard Health Publishing 2019). Specifically, sentiment analysis is commonly used in social media text processing in order to determine the opinions and feelings of social media users such as the recognition of words such as “guilt,” “impossible,” “hopeless,” “my fault”, “happy”, “wonderful,” etc (Nikfarjam et al. 2012). In this study, we will be performing sentiment analysis on tweets from the year 2020 in order to determine the existence of depression in the users. Similar work has been done in a study about how reddit users show sentiment online during the COVID-19 pandemic (Low et al. 2020). The unique approach for this study is the fact that all the tweets being analyzed are from 2020 and the COVID-19 global pandemic of 2020 and 2021 provides a special view of mental health and how people cope with global trauma and stress. Not only this, but we will be expanding our sentiment analysis to include not only emotion-words but also words associated with events and topics that may cause depression. These topic words will be largely based on researched causes of depression (Harvard Health Publishing 2019).

There are many of these causes of depression defined by the Mayo Clinic (2018) and of those, an important factor is the existence of stressful life events (Harvard Health Publishing 2019). In a previous work, we showed how the existence of a stressful life event in someone’s life is a strong predictor of the existence of depressive disorder in that person (Breeden 2021) . For this reason, we will be focusing on sentiment regarding not only emotions and feelings key words, but terms involved in stressful life events such as abuse, trauma, *global pandemics,* persisting health issues, etc.

Overall there has been a host of studies involving mental health data and NLP sentiment analysis separately, but we would like to investigate the specifics of how this data is affected by a global experience of a life-threatening, inhibiting, pandemic. Additionally, in a case where the majority of people in the world are focusing on the same event, the social media sphere may show less diversity in topic discussion. This provides us with a great way to test the performance of these typical NLP methods in sentiment analysis when the distribution of sentiment may be irregular or skewed.

Data

The data used for this study is a dataset of dutch tweets from the year of 2020. This dataset spans January 2020 through September 2020 and contains information regarding the location, original text, translated text (from Dutch to English), username, etc of all the tweets. In particular, for this study we will be focusing on the translated text in order to determine sentiment and topics that may relate to depression. Additionally, this dataset incorporates a sentiment score gleaned from each individual tweet by the “pattern” package for python. This scoring method scores the sentiment of each tweet from -1.0 to 1.0 ranging from extremely negative sentiment to extremely positive sentiment. This sentiment pattern score will be used in our classification labeling method for our data to indicate levels of negative and positive sentiment.

The data was sourced from the Twitter API and published for public use by Kaggle. The dataset contains a total of 271342 tweets from 65843 users as of April 2021 and all the users in this dataset were located in the Netherlands.

Methods

Data preprocessing

One of the most important steps in NLP is the cleaning and preprocessing of the text data. In this case, it is even more important because we are working with tweets which include symbols, emoticons, screen names, etc. For this study, we started by downloading all the separate datasets of tweets (separated by month) and combined them into one master dataset containing all data from January 2020 to September 2020. Next we dropped all instances where we had missing information for the translated text of the tweets.

We performed tokenization, stemming, and stop word elimination on the translated text and also eliminated the user handles to cut down on noise. We then added a new column to the dataset containing this new processed and tokenized text.

Exploratory Data Analysis (EDA)

After performing this text cleaning, we began some EDA in order to investigate prominent terms, relationships, and general distribution information about the dataset. This was an extremely crucial step in our analyses as this informed our labeling method for the sentiment classification of these tweets. We started by counting the most commonly used words throughout all the processed text and creating a global vocabulary for the text. The most common 20 words across the entire dataset are shown in Figure 1 (note that these words are not stemmed or lemmatized for ease of viewing for the reader).

From this vocabulary, we also determined the most prominent words found in the tweets with a sentiment pattern value of between (-1.0, -0.6) — strongly negative — and a value of between (0.6, 1.0) — strongly positive. The resulting words and their frequencies are shown in Figures 2 and 3.

WORD	FREQUENCY
corona	57514
people	21702
coronavirus	19282
still	9992
virus	8941
covid	8104
lockdown	8039
government	6634
china	6105
first	5870
pandemic	5780
really	5638
netherlands	5565
think	5547
already	5282
measures	5071
number	4795
health	4254
infected	4146
crisis	4085

Figure 1. 20 most common words
across the entire dataset

WORD	FREQUENCY	PERCENTAGE_TOTAL
corona	14776	0.246070
people	7069	0.117722
coronavirus	4094	0.068179
still	2967	0.049410
infected	2940	0.048961
virus	2337	0.038919
covid	1841	0.030659
lockdown	1819	0.030292
government	1760	0.029310
measures	1571	0.026162
china	1526	0.025413
really	1498	0.024947
netherlands	1451	0.024164
already	1434	0.023881
think	1384	0.023048
first	1364	0.022715
number	1270	0.021150
pandemic	1121	0.018668
without	1107	0.018435
nothing	1095	0.018235

Figure 2. Most common words
found in “strongly negative” tweets.

WORD	FREQUENCY	PERCENTAGE_TOTAL
corona	21573	0.230365
people	8285	0.088471
coronavirus	5182	0.055335
still	4125	0.044048
really	3529	0.037684
lockdown	3408	0.036392
virus	3031	0.032366
government	2631	0.028095
think	2525	0.026963
covid	2481	0.026493
measures	2159	0.023055
netherlands	2145	0.022905
number	2120	0.022638
already	2089	0.022307
positive	2048	0.021869
first	1988	0.021229
pandemic	1918	0.020481
china	1778	0.018986
infections	1772	0.018922
crisis	1634	0.017449

Figure 3. Most common words
found in “strongly positive” tweets.

As we can see from the Figures 1, 2, and 3, a major subject of the tweets in our dataset is the COVID-19 pandemic with words like “Corona,” “Virus,” “Covid,” and “pandemic.” This will be an important fact to remark when analyzing the performance of our algorithms. We can see that we are handling a very specific dataset lacking a large amount of topic diversity.

Because of the observations we made about the words found across different sentiment scores, we added a column indicating the classification labels for our sentiment classification model. For this label, we binned the “sentiment_pattern” column into 5 bins in order to follow the pattern found in our EDA when looking at the prominent stressful components and depressive tendencies across sentiment scores. We were informed by key words such as “nervous,” “anxious,” “worried,” “ill” to indicate the existence of “stressful life events and depressive feelings” (Harvard Health Publishing 2019). We labeled these bins as “strongly negative,” “moderately negative,” “indifferent,” “moderately positive,” and “strongly positive.” As is shown in Figures 2 and 3, the existence of words such as “virus,” “covid,” and “pandemic” is much more prominent in the negative tweets than in the positive tweets. From this and from our knowledge that stressful life events may predict the existence of depression (Breeden 2021), we can see that these “strongly negative” tweets may indicate depressive tendencies in users.

Another important fact to remember about our data is that this data is limited to twitter users from the Netherlands and therefore cannot reflect the sentiment of the entire world toward the COVID-19 pandemic.

Algorithms

After performing all the necessary data preprocessing, we separated our tokenized text into training and validation sets and padded our vectorized text.

For this study, we investigated performance of both a simple RNN model and an LSTM model for the sentiment analysis and classification of these tweets. We designated our labels as the five aforementioned classifications: strongly negative, moderately negative, indifferent, moderately positive, and strongly positive using the values of 0-4 respectively. Recall that these labels were informed by the sentiment polarity scores from the pattern library and the frequency of words referring to depression and possible traumatic events (like a global pandemic). The algorithm structures are shown in Figures 4 and 5 for both the RNN model and the LSTM model.

```
Model: "sequential"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
simple_rnn (SimpleRNN)       (None, 29, 20)            440
_________________________________________________________________
dropout (Dropout)            (None, 29, 20)            0
_________________________________________________________________
flatten (Flatten)            (None, 580)               0
_________________________________________________________________
dense (Dense)                (None, 5)                 2905
=================================================================
Total params: 3,345
Trainable params: 3,345
Non-trainable params: 0
```

Figure 4. Simple RNN model architecture.

```
Model: "sequential_2"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
lstm (LSTM)                  (None, 29, 20)            1760
_________________________________________________________________
dropout_2 (Dropout)          (None, 29, 20)            0
_________________________________________________________________
flatten_2 (Flatten)          (None, 580)               0
_________________________________________________________________
dense_2 (Dense)              (None, 5)                 2905
=================================================================
Total params: 4,665
Trainable params: 4,665
Non-trainable params: 0
```

Figure 5. LSTM model architecture.

Tuning

In natural language processing, there are a lot of moving parts which can lead to many different areas that require tuning. On top of the typical hyper-parameter tuning requirements of neural networks, this project requires tuning when approaching the input data as well. We have run all of our algorithmic variations on versions of the data that are stemmed and lemmatized and also on the versions of the data that maintain the full words (leaving out the stemming and lemmatization). We decided to run both of these versions of the algorithms in order to capture the nuances between topic discussion since we were dealing with such a non-diverse dataset in regard to subject of the tweets. The results of all the variations of our algorithms, input data, and label binning are shown in the following section.

Results

As is shown in the following figures, we can see that we ran multiple variations of our algorithms in order to determine best performance. Fig 6 shows our variations of performance with input data that has not undergone stemming and lemmatization. Fig 7 shows our variations of algorithms performance with input data that has undergone stemming and lemmatization. As you can see from our accuracy measures, we seem to achieve very similar accuracy between training and validation sets, meaning that we are neither overfitting or underfitting our data. Additionally, across all our variations of our data we can see that we seem to be achieving around 70% accuracy on our validation sets. This means that our algorithms seem to be able to correctly classify the sentiment category of these tweets about 70% of the time.

	Model_type	Num_epochs	Optimizer	Dropout	Loss_func	Number_bins	Training_acc	Validation_acc
0		0	na	0.0	na	0	0.000000	0.000000
1	RNN	5	rmsprop	0.2	categorical_crossentropy	5	0.696290	0.697934
2	RNN	5	adam	0.2	categorical_crossentropy	5	0.696340	0.697919
3	RNN	5	rmsprop	0.1	categorical_crossentropy	5	0.696315	0.697919
4	RNN	5	adam	0.2	categorical_crossentropy	5	0.696433	0.697919
5	LSTM	5	rmsprop	0.2	categorical_crossentropy	5	0.696685	0.697919
6	LSTM	5	rmsprop	0.1	categorical_crossentropy	5	0.696295	0.697919

Figure 6. Algorithm performance when
input test is not stemmed and lemmatized.

	Model_type	Num_epochs	Optimizer	Dropout	Loss_func	Number_bins	Training_acc	Validation_acc
0		0	na	0.0	na	0	0.000000	0.000000
1	RNN	5	rmsprop	0.2	categorical_crossentropy	5	0.696374	0.697919
2	RNN	5	adam	0.2	categorical_crossentropy	5	0.696364	0.697919
3	RNN	5	rmsprop	0.1	categorical_crossentropy	5	0.696084	0.697919
4	RNN	5	adam	0.2	categorical_crossentropy	5	0.696433	0.697919
5	LSTM	5	rmsprop	0.2	categorical_crossentropy	5	0.696768	0.697993
6	LSTM	5	rmsprop	0.1	categorical_crossentropy	5	0.696670	0.697949

Figure 7. Algorithm performance when
input test is stemmed and lemmatized.

Analysis and Interpretation

As is shown in the above results section, we do not have much variation in performance between algorithms, though we do see a slightly better performance in validation accuracy when running our LSTM model with a dropout rate of 0.2. Additionally, we noted that the algorithmic performance across epochs does not vary much is is shown in Figure 8. The performance of all our algorithmic variations followed almost exactly the same pattern as is shown in Figure 8. This shows us that the algorithmic performance is pretty static, meaning that the variation in performance is likely to do with the input data.

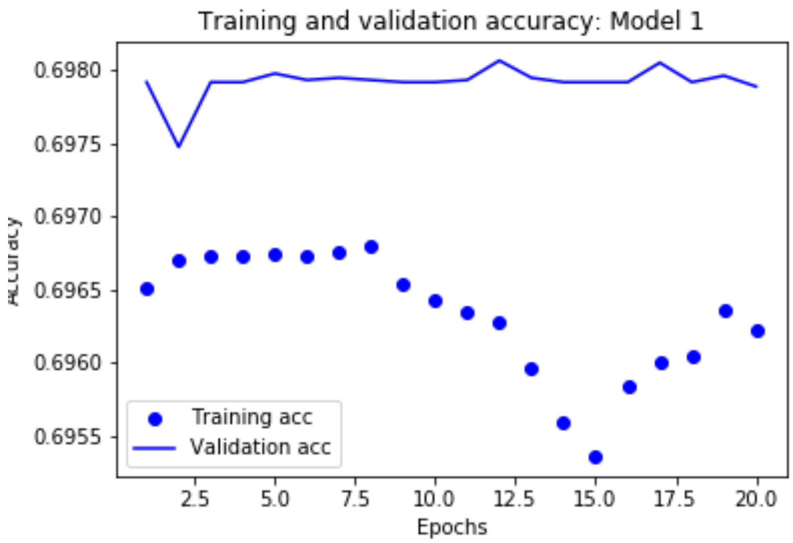

Figure 8. Algorithmic accuracy over time
for RNN model at index 1 in Figure 7.

When we interpret the data we are working with, we can recall that many of the words found in the negative tweets and the positive tweets are the same. Subjects surround COVID-19 permeate the entirety of our dataset, creating a more nuanced and difficult situation for our neural networks to work through. We are asking our neural networks to classify the sentiment of these tweets in an environment where global sentiment was very confused and largely focused on one major event. This creates a very unusual window of data that we have to work with. As we can see from the results of our algorithms, it is feasible to achieve around a 70% accuracy in classification using RNN and LSTM models even with very specific and nuanced input data.

Conclusions

Overall we can conclude from this type of sentiment analysis that global trauma creates a very interesting and nuanced social media environment. These sort of events that take over the world and affect the majority of people create an interesting situation for sentiment analysis. It is not uncommon to find instances of positive and negative sentiment in discussing the global traumatic events. We can see from our data and our algorithmic performance that when the world is mostly discussing one major event, the typical NLP algorithms used in sentiment analysis perform acceptably but not extremely well. In particular, we can conclude from these analyses that in regard to our sentiment classifications, an algorithm like this could be used in the future to flag users whose sentiment category falls in “strongly negative” in order to offer resources or support services for possible depressive disorders.

Future Work and Possible Limitations

In the future, it would be very interesting to expand the data fed into these algorithms. It would be a fascinating project to attempt to get a more global look at the social media sentiment toward COVID-19 by using tweets spanning the entire globe.

Another interesting next step could be to use social media data from live broadcast transcripts rather than written tweets. If we were able to use transcripts from Instagram Live broadcasts or TikTok Live transcripts, we may be able to capture more realistic and truthful sentiment that is less filtered and refined.

In order to work on the accuracy of these algorithms, it would be interesting to see how this algorithm works with data gathered from other traumatic events or events where the majority of social media was posting about the same subject. Social media trends and focused global events create such a specific environment for NLP that requires a more nuanced model in order to parse through different user sentiments. Expanding this model to handle difference sorts of specific environments would be a great next step.

Overall we can see from this research that RNN and LSTM models are able to achieve relatively acceptable accuracy in sentiment analysis even when handling extremely focused and specific data. We can see that even during unprecedented events, our algorithms can handle the nuance and uncertainty and are able to provide us with better-than-chance accuracy when classifying sentiment of text data. Moving forward, we will be able to trust the results of these algorithms more even when handling uncommonly non-diverse data.

Bibliography

Azadeh Nikfarjam, Ehsan Emadzadeh, and Graciela Gonzalez. 2012 "A Hybrid System for Emotion Extraction from Suicide Notes." *Biomedical Informatics Insights 2012*, no. Suppl. 1: 165–74. https://doi.org/10.4137/BII.S8981.

Breeden, Kira 2021 "Neural Network Classification of Depressive Disorder Based on Prior Stressful Life Events" Essay, Northwestern University

"COVID-19 Is Changing How, Why and How Much We're Using Social Media." March 1, 2021. Digital Commerce 360. Vertical Web Media LLC. https://www.digitalcommerce360.com/2020/09/16/covid-19-is-changing-how-why-and-how-much-were-using-social-media/.

"Depression (Major Depressive Disorder)." February 3, 2018. Mayo Clinic. Mayo Foundation for Medical Education and Research. https://www.mayoclinic.org/diseases-conditions/depression/symptoms-causes/syc-20356007.

Lane, Hobson, Cole Howard, Hannes Max Hapke, and Arwen Griffioen. 2019 *Natural Language Processing in Action Understanding, Analyzing, and Generating Text with Python*. Shelter Island, NY: Manning Publications Company.

Lin, Chenhao, Pengwei Hu, Hui Su, Shaochun Li, Jing Mei, Jie Zhou, and Henry Leung. 2020. "Sensemood: Depression detection on social media." In *Proceedings of the 2020 International Conference on Multimedia Retrieval.* 407-411.

Low, Daniel M, Laurie Rumker, Tanya Talkar, John Torous, Guillermo Cecchi, and Satrajit S Ghosh. 2020. "Natural Language Processing Reveals Vulnerable Mental Health Support Groups and Heightened Health Anxiety on Reddit During COVID-19: Observational Study." *Journal of Medical Internet Research 22,* no.10: e22635–e22635. https://doi.org

10.2196/22635.

"The Evolution of Social Media: How Did It Begin and Where Could It Go Next?" March 3, 2021. Maryville Online. https://online.maryville.edu/blog/evolution-social-media/.

"What Causes Depression?" June 24, 2019. Harvard Health Publishing. https://www.health.harvard.edu/mind-and-mood/what-causes-depression.