ML AND DATA SCIENCE

[COMPUTATIONAL LINGUISTICS]

MENTAL HEALTH
-AND -
SOCIAL MEDIA

PREDICTING DEPRESSION AND ANXIETY VIA SOCIAL MEDIA TEXTS

Nilesh Popli

ABSTRACT

Sentiment analysis has emerged as an essential system that combines NLP (natural language processing), machine learning and computational linguistics techniques in studying and analysing sentiments within a sentence or phrase. It has widespread applications, for instance, in predicting public preferences, brand monitoring, social media monitoring and market research.
It is estimated that 80% of the world's data is unstructured. Vast volumes of text data are created every day as emails, chats, social media conversations, surveys, articles, etc., but they are hard to analyse, understand, and sort through. Sentiment analysis, however, helps businesses extract meaning from unstructured text by automatically understanding, processing, and tagging it.

In this project, we have analysed how sentiments are generally expressed in our natural language and have written down our findings in the attached excel spreadsheet on the Github repo.

This has been compared and analysed our manual output against the sentiment expressed as output by the analyser and have pointed out some shortcomings and limitations with the analyser's output, along with some observations. The following report summarises the findings in accordance with the given questions.

REPORT

I.

Manually analyse the output of the sentences and check where the tool has failed and why. Give a couple of examples to show your point.

A few instances where the analyzer's output contradicted the sentiment obtained manually-

- '@LOVELESS101 oh you can? Sweeeeeet.'

  Speaker expresses positive excitement by using the word sweet as sweeeeeet. Since the spelling is incorrect, the analyser fails in recognising the positive word sweet and in the absence of positive words reports the sentence as negative with a score of around 0.756. The analyser has not accounted in the various versions of writing the word sweet.

- '@Antwoinne Hey babe, nothing much tryin to see what imma do at work today lol, look like the load isnt so bad.'

  The speaker expresses positive sentiments by expressing the work isn't so bad. The analyser may have misinterpreted the sentence as negative by identifying bad as a negative word and has failed in accounting that the negative (here isn't) of a negative word (here bad) gives a net positive sentiment (here isn't so bad which implies good).

- ‘watching gossip girl...10 episodes late.’

  Without context, it seems unclear what the sentiment of the user is, and he is just stating a fact. The analyser might on seeing the word late inferred a negative meaning because often late is used in a sense as being late. So the same word is often positive or negative depending upon the context in a sentence.

- ‘The kids are running me ragged hard work today. MrsB is off having a lovely time.’

  Speaker expresses negative sentiments of looking after kids while another person is having a lovely time. The analyser interprets the strong positive word lovely and gives the sentence a positive sentiment with a score of around 0.998, but it fails in understanding that another person and not the speaker is having a lovely time from the context. The analyser should focus on adjectives related to the speaker, such as ragged and hard as negative sentiments instead focuses on the positive sentiment lovely, which is actually associated with another person and not the speaker.

- ‘I'm write a song about you and then make a beat to ya words then later we can turn it into a hit. .lett me stoppp’

  Speaker expresses positive excitement in making a hit song. Here the word hit has been used in a positive sense, but it seems like the analyser might have interpreted the negative meaning of hit. An important observation is that often the same word might express different sentiments based on the context of its usage in the sentence.

- ‘It's hot & i can't sleep’

  Speaker expresses negative frustration at a negative situation. Analyser might have interpreted a lack of negative words as a positive sentence and has not deduced inference from the context that a hot condition often implies an unpleasant sleep which is a negative situation.

- ‘I'm starting my class today, but I'm also going to the high school afterwards to do some much needed practicing on a marimba’

  Speaker just expresses a neutral fact that is the speaker's schedule and lacks any sentiment as such. The phrase much needed, which is often used in a negative context, might have left the analyser to believe that the sentence is negative. The low sentence score of around 0.53, one of the lowest in the set, points towards this ambiguity.

- ‘ Aww its my Nanas song’

  Speaker expresses positive sentiments of nostalgia which the analyser has failed to correctly interpret probably due to the lack of positive word and also hasn’t accounted for the various spelling versions of aww which is often used in a positive sense. Also, analyser fails to interpret meaning from the context of the sentence.

II.

Does entire sentential polarity depend on a single word of a sentence? Answer with examples from the corpus.

No, the entire sentential polarity does not necessarily depend on a single word of the sentence. Below are multiple examples proving this statement.

- ‘@Antwoinne Hey babe, nothing much tryin to see what imma do at work today lol, look like the load isnt so <n>bad</n>.’

Speaker expresses positive sentiments by expressing the work isn’t so bad. The analyser may have misinterpreted the sentence as negative by identifying bad as a negative word and has failed in accounting that the negative (here isn't) of a negative word (here bad) gives a net positive sentiment (here isn't so bad which implies good).

III.

Other observations

- ‘I'm write a song about you and then make a beat to ya words then later we can turn it into a hit. .lett me stoppp’
  In the above example, the speaker expresses positive excitement in making a hit song. Here the word hit has been used in a positive sense, but it seems like the analyser might have interpreted the negative meaning of hit. An important observation is that often the same word might express different sentiments based on the context of its usage in the sentence.

- ‘@LOVELESS101 oh you can? Sweeeeeet.’
  ‘ughh i sooo sleepy.. bck 2 work i go’
  ‘Aww its my Nanas song’
  In the above examples, we observe that emotions are often expressed by writing words such as ‘aww’ and ‘ughh’. Furthermore, the emphasis is often laid on emotions by writing words such as sweet as ‘sweeeeeet’. The analyser often fails in accounting all the different versions and spellings of writing these words and often gives wrong sentiments as outputs.

- ‘http://twitpic.com/6cwst - Session of photoss! ‘
  From the words alone, there is no positive indication might be the reason the analyser infers it as a negative sentence. The sentence itself is neutral, but the many exclamation marks after the word photos make it a positive sentence as it expresses the speaker's excitement. An important observation here is that punctuation such as exclamation marks often contributes towards conveying the sentiment of a sentence.

- Emojis in text format such as ‘>.<’ or ‘:)’ often very easily convey sentiments of a sentence and should be given special attention while analysing sentence sentiment and emotions.

REFERENCES

For information regarding workings and applications of sentiment analysis, we have researched from the below sites:

- https://monkeylearn.com/sentiment-analysis/
- https://en.wikipedia.org/wiki/Sentiment_analysis#:~:text=Sentiment%20analysis%20(also%20known%20as,affective%20states%20and%20subjective%20information.