Metaverse, an Analysis

Cheolyeon BYUN
1M170020

Data and Social Media Analysis
byuncheolyeon@akane.waseda.jp

2023/01/29

Abstract

This paper shows findings that were made through the process of sentiment analysis, explores the results as well as the limitations. The subject matter is the public's reaction in relation to Mark Zuckerberg and his endeavors into making the metaverse mainstream.

1 Introduction

The metaverse has piqued entrepreneur Mark Zuckerberg's interest, and ever since Facebook changed it's name to Meta in October of 2021, Mark has started obsessing about the metaverse, and saw it as the next big thing in the world of tech. However, the public's response was very underwhelming judging from responses found on social media platforms like twitter and youtube, though there weren't a lot of credible and reliable news sources that can be cited to back up this fact. PICCHI (2022) provides one concrete fact about Meta, which is that last October in 2022, Meta's stock prices plummeted by 67 percent compared to prices from a year ago. Even after taking into account a huge hit on stock prices with every big tech companies in general, the drop in Meta stock prices were almost twice as much as other tech giants.

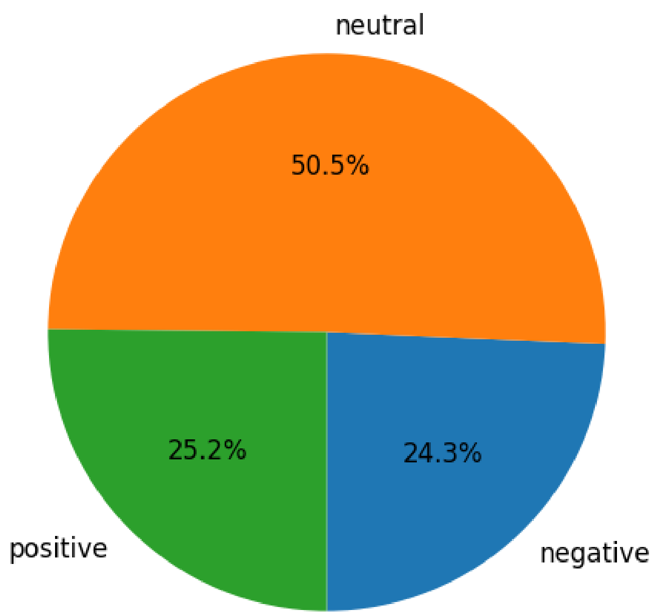
2 Data and Method

Comments were scraped from youtube using a python library named youtube comment scraper made by DataKund that utilizes browser automation to fetch youtube comments. Comments from Meta's official youtube channel doesn't exist because the comment sections there are disabled all together, which meant that I had no choice but to get comments from other famous youtuber's videos that addresses the metaverse and mark zuckerberg. The video from which the comment was scraped is titled 'Metaverse is Dead'. Then, the necessary pre-processing of data was done. To ensure that most of the relevant data stays in tact, the following measures were taken into account. First, emoticons and emojis were converted to text that represents the maening of it, this way the laughing emoji will just turn into the word 'laugh', instead of it disappearing after filtering out irrelevant characters. Regex was used to remove characters that were not alpha numerical characters. The text file that was output after scraping comments were separated by new line for each sentence. Sentences that had less than 3 words in them were removed, as those almost always consisted of usernames or text from youtube's interface itself that won't give useful information. Stopwords were removed and words were lemmatized and converted into lowercase. The first bit of sentiment analysis was done with a lexicon based approach using textblob, then a second one was done with a machine based approach that uses BERT, each result saved in a csv file.

	comments	TextBlob_polarity	TextBlob_subjectivity	TextBlob_label	BERT_label	BERT_score
0	mark zuckerberg meta facebook instagram and th...	0.000000	0.000000	neutral	neutral	0.617330
1	winning the personal device of the future	0.166667	0.391667	neutral	positive	0.869627
2	facebooks collapse the metaverse crusade	0.000000	0.000000	neutral	negative	0.778830
3	this year everything is changing facebook sto...	0.079654	0.267100	neutral	negative	0.935085
4	mark zuckerberg meta facebook instagram and th...	0.000000	0.000000	neutral	neutral	0.617330
...	...	...	...	...	...	...
2169	tik tok sounds gay	0.416667	0.583333	positive	neutral	0.549268
2170	i liked my gt	0.600000	0.800000	positive	neutral	0.681989
2171	anybody else annoyed by his voice	-0.400000	0.800000	negative	negative	0.887189
2172	way too homophobic	0.000000	0.000000	neutral	negative	0.905911
2173	your audio edits are jarring	0.000000	0.000000	neutral	negative	0.957182

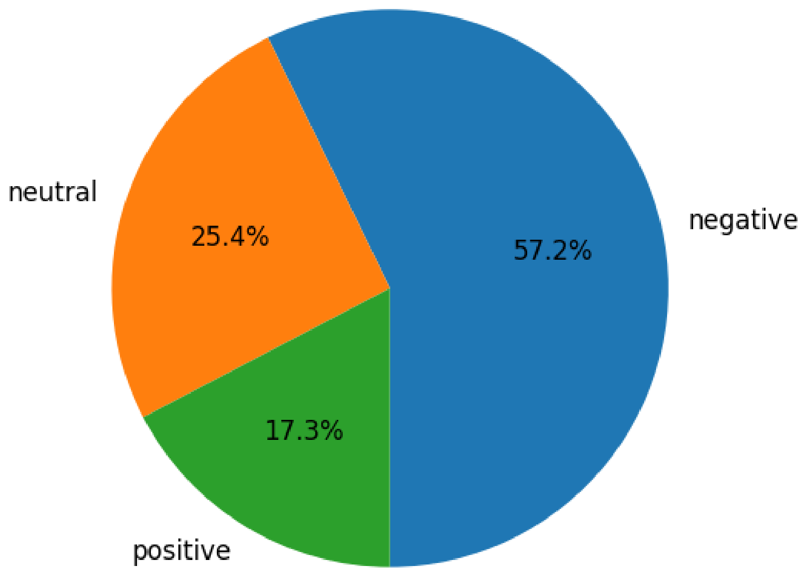
3 Results

Here are the results in a pie chart. Two pie charts were made, one created with the labels made from TextBlob data, and another created with labels from BERT. To be precise, instead of using BERT a less CPU heavy DistilledBERT model was used. Textblob sentiment scores were labeled basically on a scale between -1 and 1. The labels of positive, negative, and netural were made with conditions I have arbitrarily created, Texts labeled positive needed a polarity score of over 0.2. Texts labeled negative had a score of under 0, and in between were labeled neutral. These random looking thresholds were made after already looking at the results with different numbers, where the data labeled neutral were too abundant, and thus the only data with a polarity score between 0.2 and 0 were labeled neutral. With data using BERT, labels for positive and negative were already pre-built into the module so that those labels were given right away. I had to take one extra step to create a neutral label in order to compare it with the other data, though the criteria was created arbritarily again. This time, BERT scores under 0.7 were labeled neutral. BERT scores does not go under a zero, and was always between 0 and 1, that only indicated how positive or how negative each sentence was. The score 0.7 was decided again after seeing the results already, where this time under a 0.5 showed very little data labeled neutral.



Pie Chart (TextBlob)

This here is a pie chart made from TextBlob labels. Despite the fact that only data with sentiment polarity score between 0 and 0.2 were labeled neutral, it still took up more than 50% of all data.

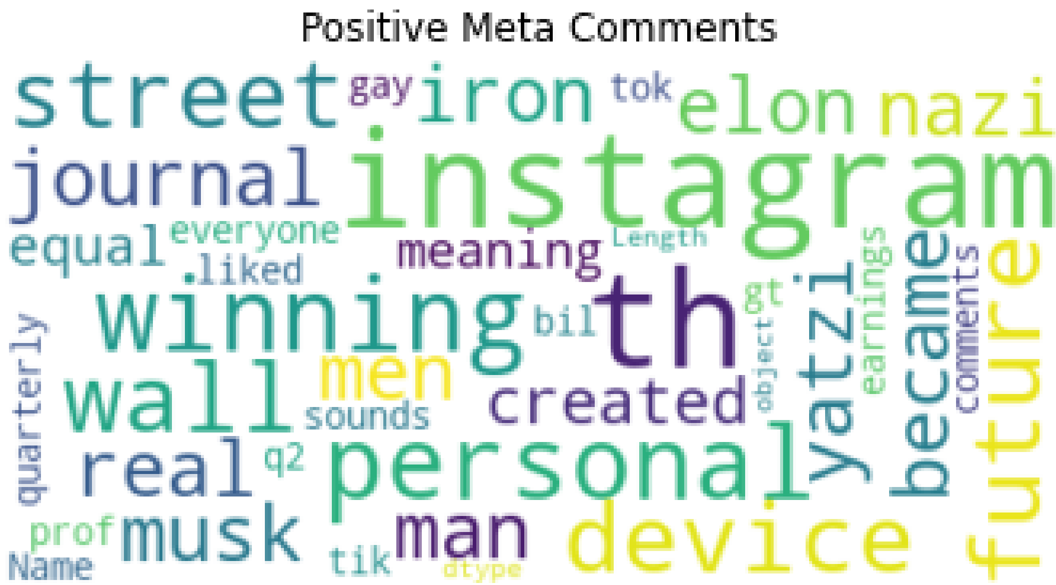


Pie Chart (BERT)

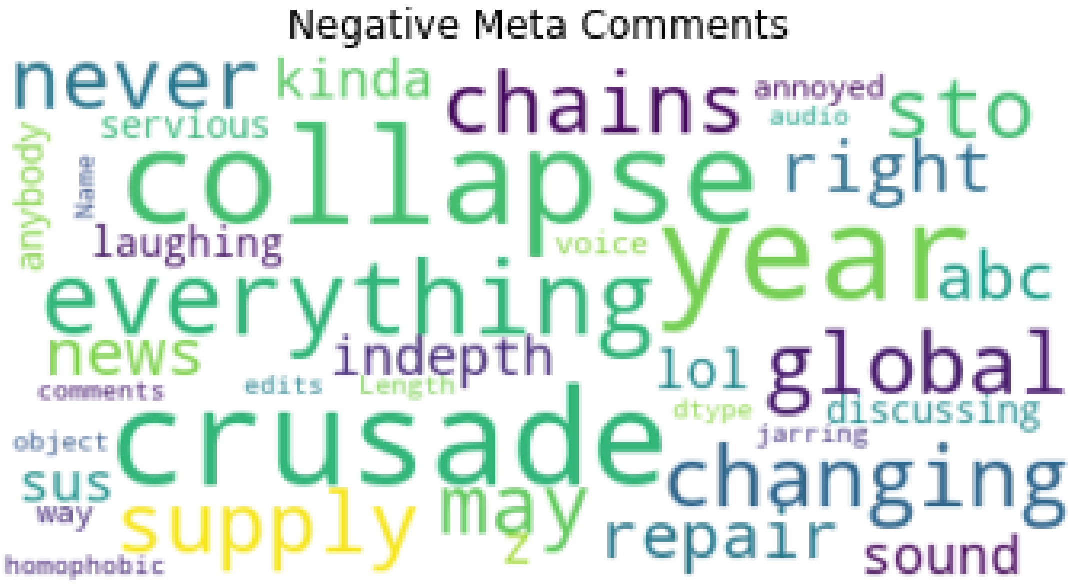
This pie chart was made from BERT labels. In here data with a BERT sentiment score of lower than 0.7 were all labeled neutral. And yet, negative comments, which were between only a 0.7 and a 1 in BERT sentiment score, consisted of more than 50% out of 2174 sentences.

4 Conclusion and Limitations

The BERT model's accuracy and F1 score was trained using IMDB dataset of movie reviews, and both scored a 0.873, which is relatively high. Though it is very tempting to conclude that a majority of negative data could directly translate to the public's reception of the metaverse, more due-diligence is required. First off, the data itself includes comments such as, 'this is a great analysis video!'. This obviously refers to the video, and subsequently, the commenter agrees with the video about the fact the metaverse is dead, as the title of the video suggests. However, the fact that it's an affirmative of the negative cannot easily be detected. The comments aren't just about trash talking meta and mark zuckerberg, they're usually a mix of complimenting the youtuber and their video, and ridiculing Mark Zuckerberg, and sometimes the comments didn't contain any useful information to begin with. Secondly, a bubble filter effect, where one is exposed to the kind of data that already confirms their beliefs because of how algorithms for content works, could be attributed to the reason why negative comments seemed dominant. Below are wordclouds of the data from BERT to give a sense of what words were actually used the most for positive comments and then for negative comments, and why the data itself might not have been the best data to conduct an analysis to begin with.



WordCloud (BERT) - Positive



WordCloud (BERT) - Negative

As evident by the wordclouds, especially the one from positive comments, the words used the most doesn't really make a whole lot of sense. Apart from the word 'winning', most other words doesn't give one a good indication why it's positive, or why it's even relevant to the metaverse and Mark Zuckerberg, such as the words 'personal', 'wall', and 'street'. One explanation would be that the word 'wall street' was used which is mentioned because Meta was losing inverstors left and right, but Meta losing investors should mean that the word wall street should have been heavly associated with a negative connotation, not the other way around. The negative wordcloud seems to suffer from the same problem, where the comments themselves seem like they don't show you words that seem germane with the issue of the metaverse failing.

5 Future Work

In the future, I'll take more time increasing the accuracy and F1 score for a better result, by using a myriad of different feature representations and identifiers for my data. Also, I'd try to get some better data used in the first place. Though I'm not entirely sure how I'd be able to get cleaner, more relevant comments, perhaps I could select a youtuber that's a tech youtuber, that covers and addresses tech related stuff more that should attract more people that are better informed and are less likely to troll with their comments by leaving irrelevant information.

6 Code

`https://github.com/OWO-hue/Metaverse-Sentiment-Analysis.git`

References

PICCHI, A. (2022). Meta's value has plunged by $700 billion. wall street calls it a train wreck.