A report on Elon Musk and Twitter: Analyzing themes and sentiments

Aniket Kale | Spring 22 | Word count: 1800

1. Introduction

Elon Musk is one of the most influential people of recent times. He is currently the richest person on the planet and frequently likes to share his opinions on Twitter. The topics that he discusses in his tweets are a variety of subjects like Cryptocurrency, politics, astronomy, technology, and more. His one tweet regarding the environmental impact of mining Bitcoin subsequently saw a drop in the prices of Bitcoin by a substantial amount. Another series of tweets on Bitcoins resulted in fluctuating prices of the same according to the following article by Vox magazine: "When Elon Musk tweets, crypto prices move" *(Molla, 2021)*. On April 4, 2022, he bought 9.2% shares in Twitter, Inc. People around the world had different kinds of reactions to this news, given the impacts of this decision. He raised a bid of $44 Billion to buy Twitter Inc. and it was accepted on April 25, 2022, becoming the owner of Twitter. Given his influence on the world, several concerns have been raised about the state of Twitter as a platform. These concerns mainly include topics like freedom of speech, expenditure of such a huge amount of money that could have been otherwise utilized in a more humanitarian manner, politics, and more. Nevertheless, Elon Musk buying Twitter has been a topic of discussion across Twitter since the news of Musk buying Twitter was announced.

2. Research Question

In this study, the primary goal is to analyze the themes that exist in the discussion of (1) Musk becoming the highest shareholder in Twitter, Inc. and (2) Musk becoming the owner of Twitter, Inc. Also, the sentiments of the reactions to these two events are to be analyzed. Additionally, the secondary goal is to analyze all of Elon Musk's Tweets to understand the themes that occur in his Tweets since he joined Twitter.

3. Method

To analyze the themes in a particular dataset, Topic modeling in Python is applied to a corpus of tweets. Topic modeling will thus highlight the most represented keywords in a particular topic These keywords can then be used to interpret the title of the deduced topics.

3.1. Data

3.1.1. Data Collection

For the purpose of this study, the reactions of Twitter users across the world are to be recorded pertaining to the two events i.e. Musk buying shares in Twitter and Musk becoming the owner of Twitter. To keep the context subjective only to these events and possibly eliminate as much noise or unrelated tweets, the following keywords were carefully chosen to scrape tweets:

- Data-set 1: Elon Musk's Tweets and Replies since 2010
    - All of Musk's Tweets were scraped excluding those which were deleted. The Tweets were only collected till April 29, 2022.

- Data-set 2: Reactions to Musk buying 9.2% stocks in Twitter (Total count: 1500)

  - Collection period: *April 4, 2022 – April 20, 2022*
  - Keywords used: *elon/musk/Twitter/shares*
- Data-set 3: Reactions to Musk buying Twitter for $44 Billion (Total count: 1200)
  - Collection period: April 25, 2022 – April 29, 2022
  - Keywords used: *elon/musk/Twitter/billion*

The Tweets were scraped iteratively using the Python library "snscrape", using combinations of the mentioned keywords. For instance, "Elon" and "stocks" were used together to obtain the Tweets. Duplicate tweets were eliminated as much as possible.

3.1.2. Data Cleaning and Pre-processing

It is important to filter common words and special characters from the Tweets to ensure that only quality and meaningful information is utilized for analysis. If common words are not removed, it can hinder the performance of Topic-modeling techniques like Latent Dirichlet Allocation (LDA). This is because LDA relies on the frequency of words that exists in the corpus, and having common words like "the", "and" and more can misrepresent the word embeddings in the corpus. Words that are commonly used in the English lexicon, like 'if', 'they', 'so' and more are not generally helpful while applying a model for analysis in Python. These were eliminated (for Topic modeling). The data cleaning process involved removing special characters, numbers, links (URLs), and mentions (to another account. Optionally, it would be needed for the "#" character to be eliminated. The procedure for removing punctuations was also required as few models rely on punctuations for Sentiment Analysis. Python libraries like NLTK, regex, SpaCy and pysentimiento were used for the cleaning process.

3.2. Data Analysis Methodology

3.2.1. Overview

First, all of Elon Musk's individual Tweets were analyzed and his emotions in the Tweets were recorded. Then, the Data-sets 2 and 3 were analyzed separately using sentiment analysis and topic modeling in order to contrast the themes and sentiments on Twitter after the two different events. In all of the analyses, the libraries "gensim" for LDA and "Top2Vec" for using BERT (Devlin et al., 2018) as a baseline model were used for the purpose of Topic modeling. For sentiment analysis, the "pysentimiento" library was used. This library includes models trained using RoBERTa (Pérez et al., 2021) as a baseline, trained on roughly 58 million Tweets, and fine-tuned for sentiment analysis. As a result, this provides an accurate sentiment analysis framework which is very well suited for the task at hand. For Topic modeling using LDA the techniques discussed in the following paper titled "*Topic modeling for social media content*" (Babanejad et al., 2016) and "*Data And Content Analysis For Social Network Using LDA Text Model*" (Zhang, 2019).

3.2.2. Analyzing Elon Musk's Tweets since 2009

Elon Musk has Tweeted roughly around 4265 times with over 11,600 replies. The distribution of Elon's Tweets over the years can be seen in Figure 1. All of these Tweets and replies were used for the Topic modeling after the cleaning and preprocessing steps. Data-set 1 was used for this analysis. Latent Dirichlet Allocation was used on the *bag-of-words* generated from this data-set. Upon analysis, the ideal number of topics was found to be 5, as they had slightly different themes amongst themselves. After training the model using LDA, a library called "pyLDAvis" was used to visually analyze the content in each of the topics. Additionally, Top2Vec was used to obtain the topics. This library automatically decides the number of topics analyzed in the corpus. For Data-set 1, the topics analyzed were around 140, which is absurdly high for the corpus. However, upon further analysis, there were only a few meaningful topics which can be interpreted based on their keywords. These findings will be discussed in the results section. Also, emotion analysis was performed to capture Musk's temperament in his replies.

3.2.3. Analysis of reactions on Twitter

The data-sets 2 and 3 contain the reactions of Twitter users to the two events as discussed above. Both data-sets are separately analyzed using the same pipeline. First, the data-sets are cleaned and preprocessed. Then the Top2Vec model is used to first detect the number of topics, then LDA is used to get the actual topic models based on the obtained number of topics. Using "pyLDAvis", the most frequent keywords in each of the topics are analyzed. For sentiment analysis, the pre-trained model from "pysentimiento" library was used to classify the reactions as "Positive", "Negative" and "Neutral".

4. Results

4.1. Results for analysis of Elon Musk's Tweets since 2009.

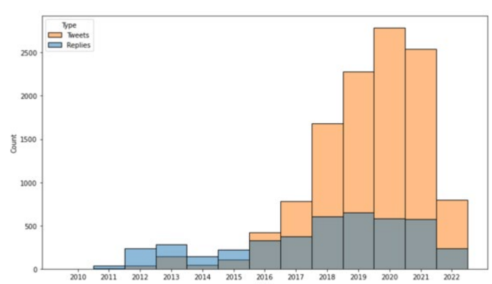

Figure 1: Distribution of Musk's Tweets by Year (Till April 2022*)

The following results were obtained as a result of LDA on Data-set 1:

Topic	***Keyword***	***Interpretation of Topic/Theme***
1	*Solar, spaceX, launch, power, voltage, beta*	*Technical; Astronomy*
2	*University, student, semester, bankrupt, engineering*	*Education*
3	*Sevices, economy, engine, raptor, thrust*	*Astronomy; Space exploration*
4	*Tesla, cars, Model, X, super chargers*	*Automobiles*
5	*Entertaining, hypocritical, voters, sigh, understood*	*Opinions; criticisms*

Emotion analysis of Elon's Tweets:

Emotion	*Percentage*
Neurtal	*91%*
Joy	*7%*
Anger/sadness/fear/surprise	*2%*

4.2. Results for reactions after Musk bought Twitter stocks

The following results were obtained as a result of LDA on Data-set 2:

Topic	*Keyword*	*Interpretation of Topic/Theme*
1	*Takeover, social, media, board, people, deal, reuses*	*Administrative*
2	*Takeover, stake, stakeholder, button*	*Administrative*
3	*CEO, shares, decentralized, radicalized, haters, triggered, Trump*	*Emotions (enraged)*
4	*Manipulated, bargain, censorship, speech, tears*	*Emotions*
5	*Purchase, platform, economics, advisers, guns*	*Economics*
6	*Chinese, races, financing, secure*	*Race*
7	*Taxpayers, democracy, Trumpesque, litigation, people, insulting*	*Politics; administration*

Figure 2: Data-set 2 Intertopic Distance Map

The following are the results of the sentiment analysis:

Sentiment	*Percentage*
Neutral	*62 %*
Positive	*8 %*
Negative	*30 %*

4.3. Results for reactions after Musk bought Twitter for $44 Billion

The following results were obtained as a result of LDA on Data-set 3:

Topic	***Keyword***	***Interpretation of Topic/Theme***
1	*Privatized, experts, game, takeover, free, speech, shareholder*	*Free speech; Privatization*
2	*Wallstreet, workers, directors, stake, media*	*People*
3	*Perspective, n*zi, tyrants, trump, censorship, free, speech*	*Emotion (enraged)*
4	*Fakenews, misinforming, propaganda, rightwing, threatening, paradoxical, cancel, system*	*Misinformation; accusation*
5	*Bring, back, American, patriot, news, jack, Disney, doge, platform*	*Patriotism*
6	*Bloomberg, CNN, nation, money*	*News outlets*

The following are the results of the sentiment analysis:

Sentiment	***Percentage***
Neutral	*46 %*
Positive	*19 %*
Negative	*35 %*

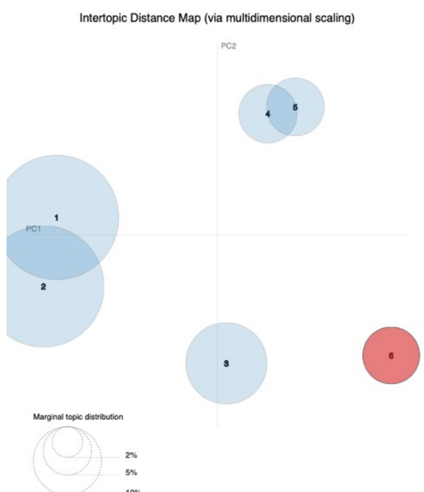

Figure 3: Data-set 3 Intertopic Distance Map

The keywords in each of the topics are selected if they are unique among topics. The interpreted topics are labeled manually based on the keywords and underlying assumptions about the nature of the Tweets, for instance, tagging a topic as “free speech” requires an understanding of the context of discussions for the given topic.

5. Conclusion:

Elon Musk’s Tweets are mainly about Tesla, SpaceX, Astronomy, and other technical terms. His replies have mostly been neutral. It can be observed that the number of positive tweets increased as Elon Musk bought Twitter as opposed to him buying 9.2% stakes in Twitter. However, on the other side, the topics like “free speech”, “censor-ship”, and “politics” increased. For data-set 2, topics 1, 2, and 3 are closely related to each other, as can be observed in the inter-topic map (Figure 2). However, for dataset3, the Topics 1, 2 and 4, 5 are closely related, as can be observed in the inter-topic map (Figure 3).

6. Limitations:

The sentiment analysis can be improved by fine-tuning the models using the obtained data-sets but it would be needed to annotate the Tweets manually. LDA requires hyper-parameter tuning to work flawlessly.

Bibliography

Molla, R. (2021, June 14). When Elon Musk tweets, crypto prices move. Retrieved April 2022, from Vox.com: https://www.vox.com/recode/2021/5/18/22441831/elon-musk-bitcoin-dogecoin-crypto-prices-tesla

Babanejad, G., Rohani, V. A., & Shayaa, S. (2016). *Topic modeling for social media content: A practical approach An Efficient Model for Processing Skyline Queries in Incomplete and Uncertain Databases View project Topic Modeling for Social Media Content: A Practical Approach*. https://www.researchgate.net/publication/321185310

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*.

Pérez, J. M., Furman, D. A., Alemany, L. A., & Luque, F. (2021). *RoBERTuito: a pre-trained language model for social media text in Spanish*. http://arxiv.org/abs/2111.09453

Zhang, L. (2019). Data and Content Analysis for Social Network Using LDA Text Model. *Journal of Physics: Conference Series*, *1213*(2). https://doi.org/10.1088/1742-6596/1213/2/022035