# Does Twitter chaos effect popularity on Billboard?

Sadhana Desai
State University of New York at Binghamton
Masters in Computer Science
sdesai13@binghamton.edu

Kasturi Vartak
State University of New York at Binghamton
Masters in Computer Science
kvartak2@binghamton.edu

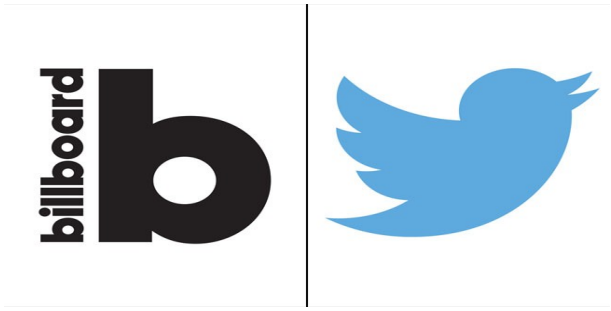



**Figure 1: Lot of Twitter chaos,will it effect popular stars?**

## ABSTRACT

People can freely express their opinions on any matter at one platform that is Social Media. Firstly, Twitter, Instagram, and Facebook are three main social networking sites where people produce a huge amount of data about users behaviors,news, new trends,etc. A large amount of user generated content is available, leading to a lot of changes in the world.Twitter undoubtedly has held its firm position among all social networking sites with an exponential number of users every year.Many studies in healthcare industry, politics, sports, and music industry are carried out investing the power of Twitter data.Music lovers are prone to interact with their favorite songs and artists through social media, which provides enormous troves of insight not on just individual song and artists but also on how music consumers perceive any song.This chaos can help to get the information on the next week's popular artists or songs.Findings of this study are beneficial to the music industry to discover song performance by real-live update trends on social media in order to propose an appropriate strategy for hit and non-hit songs.

## KEYWORDS

Twitter, Datasets, Sentiment Analysis, Artists, tweets, Retweets, Followers, Mentions, Hashtags, Songs, Billboard charts

## 1 INTRODUCTION

Nowadays, in our booming era, the freedom of expression on the internet has been simplified than ever. The widespread adoption of social media has empowered users to freely express themselves and engage in far-flung conversations with people from all over the world. Given the significance of social media, understanding user-generated content provides a rich source for management and strategists in various industries. It is no exception to the billion music industry, which has experienced an exponential growth thanks to the introduction of streaming music over the last ten years. The fast growth of technology has transformed traditional music platforms into online music services where everyone is brought closer together. Social media platforms are where music audiences naturally congregate, building their own communities, and sharing experiences of songs and artists that subsequently is a catalyst to encourage users' engagement to the music industry.
Twitter remains as the best indicator of the broader pulse of the world and overview picture of what is happening within.It has become one of the outlet for people talking about music.Given a considerable amount of music-related users' behaviors data, understanding Twitter is a potential source for the music industry in various perspectives. With all this we can say that the chaos going on in the twitter tweets, retweets, followers, mentions, hashtags can help to analyze the changes in the rankings of songs and artists of the Billboard charts.

## 2 BACKGROUND RESEARCH

The significance of social media is huge enough to see it growing in the community as a fire is nothing new. Whenever there is a new release in music industry,there's a lot of talk about it and if the artists are already popular they get even bigger fan base as more and more content is shared and retweeted about them. It is no exception to the billion music industry, which has experienced an exponential growth thanks to the introduction of streaming music over the last ten years. The fast growth of technology has transformed traditional music platforms into online music services where everyone is brought closer together.
Twitter created a very different type of environment which created the most open communication forum between fans and artists and between fellow artists, giving artists, actors and the like a chance to let their true feelings and deepest thoughts out to their huge followers. Given a considerable amount of music-related users' behaviors data, understanding Twitter is a potential source for the music industry in various perspectives. With all of this connection between Twitter and the music industry, the initial question that

comes to mind of the author is "to what extent Twitter data can effect song rankings or popularity of artists?".
Billboard charts are something which are looked up every week in USA to see the top artists and their work.The Billboard Hot 100 is the music industry standard record chart in the United States for songs, published weekly by Billboard magazine. Chart rankings are based on sales (physical and digital), radio play, and online streaming in the United States.A new chart is compiled and officially released to the public by Billboard on Tuesdays.

## 2.1 Related Work

There is a lot of research work going in the music industry.We went through a lot of work on Spotify charts and billboard charts analysis.Even Twitter was analysed to get some inputs for music industry for going through the music trends over the internet.We went through just one similar kind of research where the author wrote a thesis on predicting billboard charts.The thesis named 'Leveraging social media in the music industry'.The thesis focuses on predicting the charts for the future weeks based on the previous available charts using twitter content.This paper was more of correlating the factors between Billboard and Twitter and focused on two main research questions which were if twitter data solely depends on predicting song chart and if independent variables and song ranking are related? The author tried to correlate the factors and got through some findings to predict the charts.

# 3 DATASET DESCRIPTION

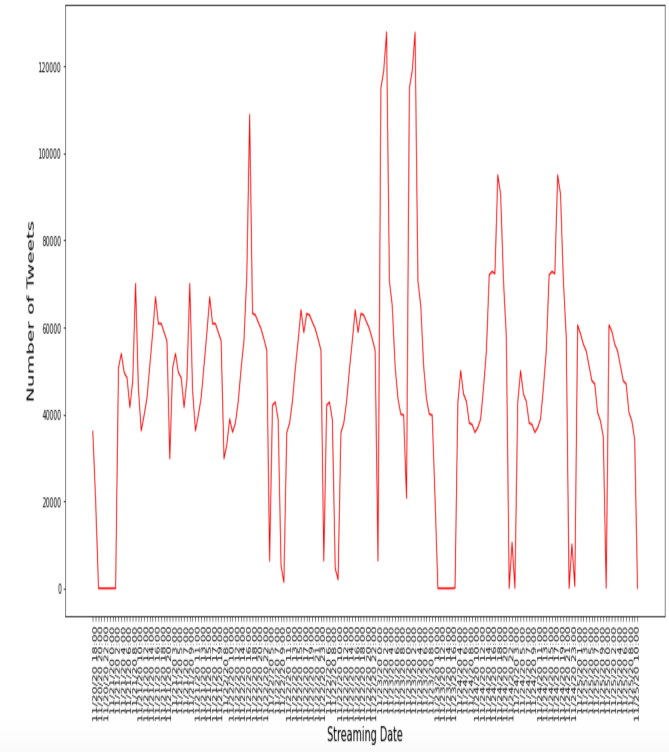
## 3.1 Twitter

Collection of stream of tweets is done using Python code.We have used the Twitter Streaming API to collect real time tweets.The resource URL used is https://api.twitter.com/2/tweets/sample/stream. We have appended several parameters at the end of the URL which were useful for the further analysis in our Project.

We planned to analyze the number of followers, mentions,retweet counts, hashtags and few factors to track the effect on Billboard charts.We have used the twitter fields available which are:

- Id - Unique identifier of the Tweet.
- Tweet - The content of the Tweet.
- CreatedAt - Creation time of the Tweet.
- Public Metrics - Engagement metrics for the Tweet at the time of the request such as count of retweets, replies and likes.
- Entities - Contains details about text that has a special meaning in a Tweet such as mentions, hashtags and other details.
- Language - Language of the Tweet, if detected by Twitter.

Below we have data of tweets collected over from Nov20 to Nov27. We have collected around 153120 tweets based on billboard filters.The graph shows the variation of titter data streaming count.




## 3.2 Billboard

There is no official API for Billboard, so we have created a script as per discussion with the professor just to get HOT-100 chart entries using https://github.com/guoguo12/billboard-charts as a reference. URL: https://www.billboard.com/charts/hot-100

As we were getting the response in HTML format, we have used BeautifulSoup as an HTML parser. Billboard API updates every Tuesday afternoon (though the time is not specified anywhere).If there is a holiday on Monday then HOT-100 chart updates on Wednesday afternoon. Thus we are running our script every Tuesday afternoon at 12 PM. After getting the chart entries, we are inserting chart entry attributes into our database given below:

- Artist: The name of the artist, as formatted on Billboard.com.
- PeakPos: the track's peak position on the chart as of the chart date, as an int (or None if the chart does not include this information).
- LastPos: The track's position on the previous week's chart, as an int (or None if the chart does not include this information).

This value is 0 if the track was not on the previous week's chart.
- Weeks: The number of weeks the track has been or was on the chart, including future dates (up until the present time).
- Rank : The track's current position on the chart.
- isNew: Whether the track is new to the chart.

We have collected Billboard data from Nov 10 till Nov 27 which is almost 3 weeks.Data collected for Billboard was directly stored in the database.After the data is collected, we stored it in a database to preserve it so we queried it as and when data was required.To do this, we used Postgres RDBMS.

[title, artist, peakPos, lastPos, weeks, rank, isNew]

['Positions', 'Ariana Grande', 1, 0, 1, 1, True]

['Forever After All', 'Luke Combs', 2, 0, 1, 2, True]

['Mood', '24kGoldn,iann dior', 1, 1, 12, 3, False]

['Laugh Now Cry Later', 'Drake,Lil Durk', 2, 3, 11, 4, False]

['Blinding Lights', 'The Weeknd', 1, 4, 48, 5, False]

['I Hope', 'Gabby Barrett,Charlie Puth', 6, 6, 44, 6, False]

['WAP', 'Cardi B,Megan Thee Stallion', 1, 2, 12, 7, False]

['Savage Love (Laxed - Siren Beat)', 'Jawsh 685,Jason Derulo', 1, 5, 20, 8, False]

['Lemonade', 'Internet Money,Gunna,Don Toliver,NAV', 9, 10, 11, 9, False]

['Holy', 'Justin Bieber,Chance The Rapper', 3, 9, 6, 10, False]

## 4 METHODOLOGY

The objective of this section is to guide through methodological decisions undertaken for this project.We have proposed the following steps in moving forward with the project:

### 4.1 Data Collection

The first step is gathering the data that we need to analyze. Prior to this paper, we have create scripts written using Python as our data collection system. We collect data from Billboard and Twitter. For Billboard we collect Billboard HOT 100 chart every Tuesday. We have used the Twitter Streaming API to collect real time tweets from https://api.twitter.com/2/tweets/sample/stream. We have appended several parameters at the end of the URL and collected only tweets in only English language which can be useful for the further analysis in our Project.

### 4.2 Data Cleaning

The next step after we have data in our hands is cleaning the data collected from Twitter. By intention we have only filtered the Twitter data in English language. Now we need to filter the this data using hashtags and mentions collected from last weeks Billboard HOT 100 chart entries. The attributes in Twitter data that we plan to extract are the video ID, title, description, author, creation date, and hashtags creation date, id, tweet, relationship between tweets, such as mentioned, retweet count etc. The extracted features are stored separately from the raw data.

### 4.3 Data Analysis

- The first step is to perform sentiment analysis on tweets collected using hashtags and mentions from Billboard HOT 100. For this, we used Vader tool.Vader performs well for the analysis of sentiments expressed in social media.

- The second step is to predict the songs solely based on twitter data using parameters like Retweet counts, number of followers, number of tweets artist has mentioned.

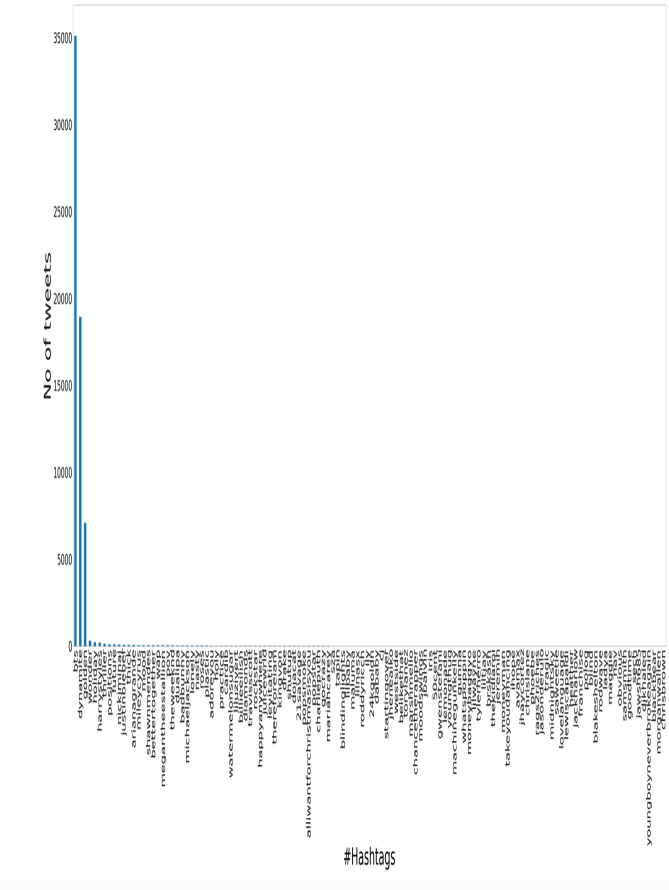
- The last step of the methodology is inferring the analysis. In this step, we have the result of the analysis. We use the result to make it meaningful. We may have to relate to other studies to draw conclusion about the result.

## 5 RESULTS

We tried to analyze our research questions and get through our topic for the popularity of the songs and artists.As mentioned above, our steps for analysis we first cleaned the data and started the tweets to be filtering out using hashtags and mentions.

### 5.1 Hashtags

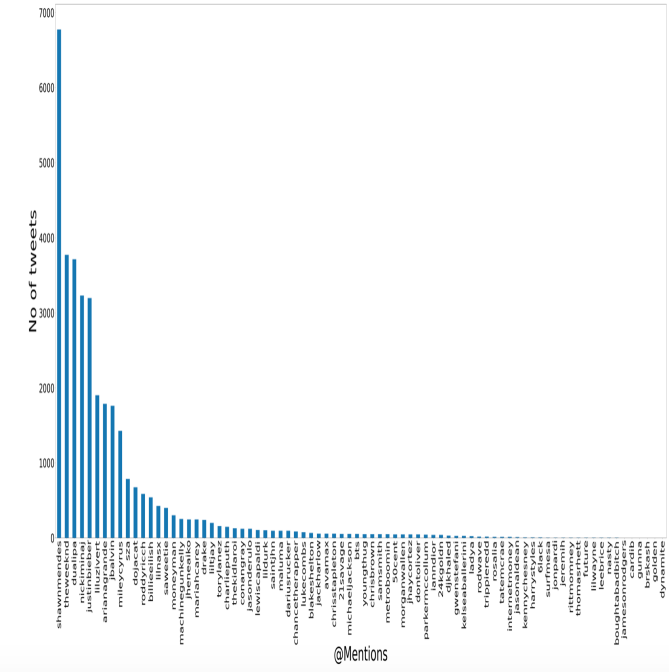
Hashtags are used a medium in social media to tag and wrap up the whole content of the tweet or posts. It is indicated by word preceding with a # symbol. It is used to index keywords or topics on twitter. Here in our dataset, we have used the Billboard top lists or the songs and artists names to clean and filter out the data for our analysis.
We have collected the hashtags which are case sensitive or as and when they can be used. Hence if the tweet has hashtages as #ArianaGrande (all together first letter uppercase) or #arianagrande (all lowercase) or #Ariana Grande (if includes space) or #ARIANAGRANDE (all uppercase). These conditions are handled so we could get the exact count of the hashtags.When we just tried searching using title without handling these conditions,we just collected 366 tweets.When we started handling the conditions, we got collected around 16391 tweets.Hence, the condition handling helped us to fetch all the tweets. The hashtags includes whole title of songs, artists names and also their group or bands as that could help us to check on the popularity factor. The below graph shows the most popular hashtags in our collected tweets.




From the above graph, we could infer that BTS has a lot of fan following base as BTS hashtag is used mostly in our tweets and their song Dynamite is on 2nd highest.The third top in the graph is Golden by Harry Styles.So here we could infer that along with songs the artists names are also being used in the tweets which could increase their popularity and reach.Hence, creating a hit for the charts.

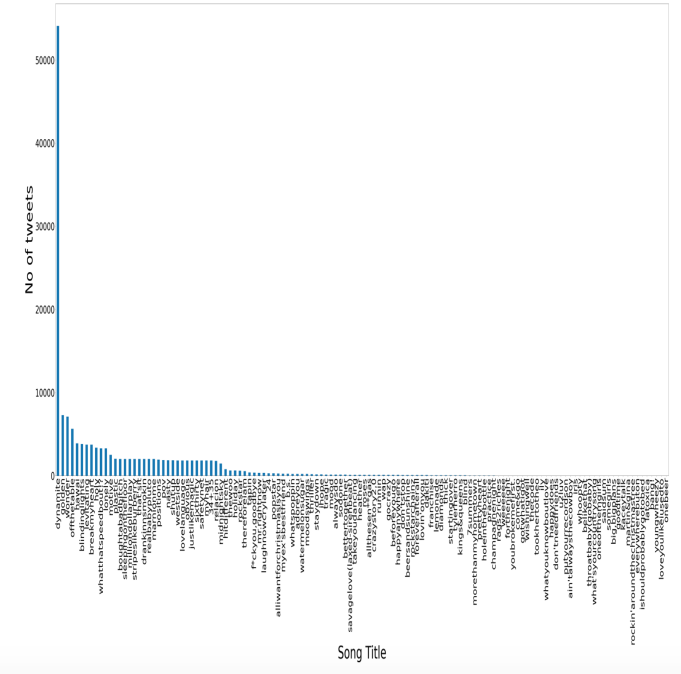

## 5.2 Mentions

Whenever there is a post about something specific or if a user on twitter want to tag up an artist, they use mentions for this purpose. Mentions in a tweet contains another username anywhere in the body of the tweet by using symbol @ to notify users that they are mentioned in the tweet. Here, there are usernames or accounts on the names of the artists, their groups or if the song reaches some popularity level then even songs have a username.
Collecting the mentions we used mention @ symbol to extracting tweets that mentioned the main Twitter account of the artist such as @edsheeran or @justinbieber.




## 5.3 Popular Songs

The songs on the charts are based on the name and the artists.Most of the times, its the artists name only works which increases the popularity of the song.Here we have analysis of combination of both songs and artists which were tagged or mentioned in tweets by the users.Tweets are mostly comprised of both the song titles and artists so here we tried to analyze the popular songs and could get their rankings from our collected data.The below graph hence then shows the popularity of a particular song on twitter:



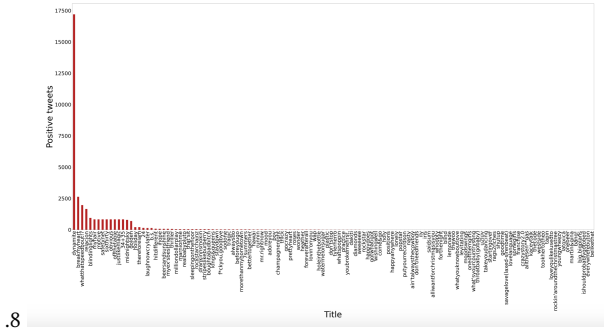




**Figure 2: Positive tweets**

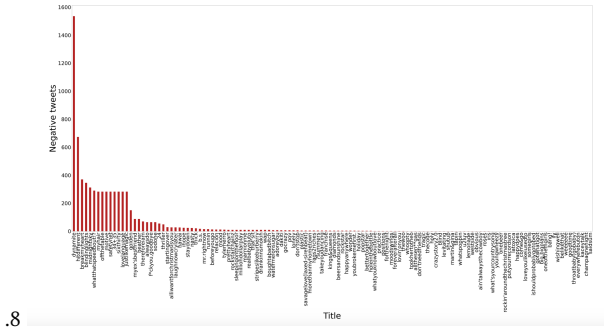



**Figure 3: Negative Tweets**

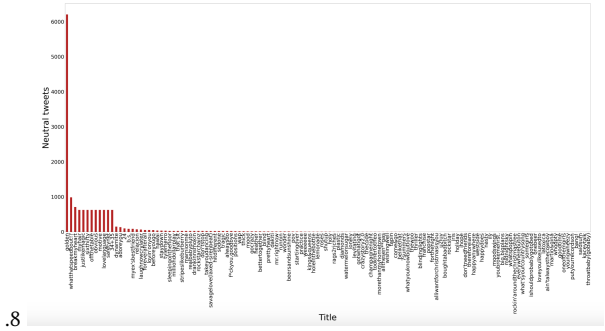



**Figure 4: Neutral Tweets**

**Figure 5: plots of sentiment analysis**

## 5.4 Sentiment Analysis

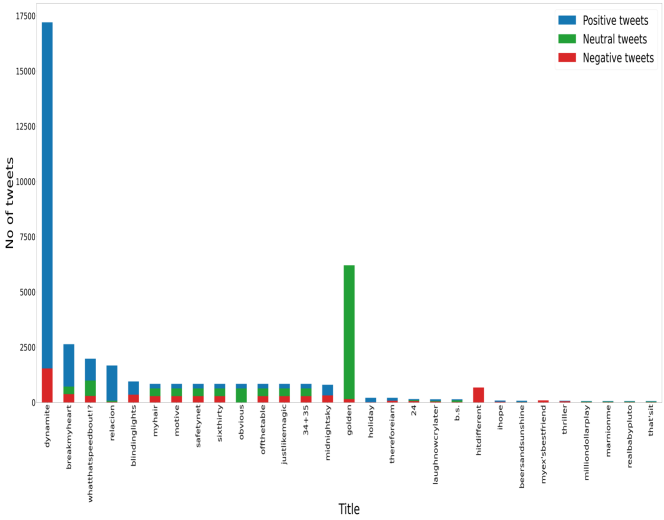
The major part of our research was to find the sentiment analysis of the tweets for the music industry.Sentiment analysis is performed for Twitter using a tool called Vader. These sentiments must be present in the form of comments, tweets, retweets, or post descriptions, and it works well on texts from other domains also. VADER sentiment model is so designed which extracts features from twitter data, formulate the sentiment score, and classifies them in positive, negative, neutral classes. Sentiment-related variables are calculated on tweets related to songs only.

In Vader Sentiment Analyzer, we have first cleaned the tweet text that is we have removed any special characters, emojies and urls and calculated the score based on the "compound" score computed to differentiated between by following below criteria:

- Positive if compound score >= 0.05
- Negative if compound score between -0.05 and 0.05
- Negative otherwise as Compound score is a metric that calculates the sum of all the lexical ratings which are normalized between -1(most negative) and +1 (most positive).

Below figures shows the positive,negative and neutral tweets about the top chart songs collected from the data.Majority of them are considered as neutral tweets as it is music most of the time it is just shared for others to know about it.The negative and positive tweets vary according to the users comments and text in the post.

The final graph below shows the combined analysis of tweets of a single song with positive,negative and neutral tweets.All three are shown in one single bar and hence we can analyze which song has more positive or negative reactions on twitter.



## 6 DISCUSSION AND CONCLUSION

We formulated three research questions to achieve our research objectives, that is to present an understanding to check effect of twitter on Billboard HOT 100 . In the process of answering the research questions, we have devised a methodology for the measurements and analysis plan which comprise of data collection, data cleaning, data analysis.We could find the sentiment analysis was effective to find out the positive,negative and neutral tweets from the tweets which we collected.Sentiment analysis helped us to infer that a popular song can have both high number of positive and negative comments.Majority of them being neutral as most tweets are just to share content in music industry.For the sentiment analysis, we saw BTS was the only exception with more number of both positive and negative tweets still being on the list on later positions. BTS is the only group with more tweets. For other songs and groups, they do have more number of tweets and positive comments and are still on the top positions.

The second question was to analyze the hashtags and mentions.As mentioned in the results, the number of hashtags and mentions show the popularity of the songs.The graphs show the songs and the artists mentioned and tagged from the dataset.The combined graph for artists and songs show that both have high number of results.Sometimes, the group being on top 30 may also get the songs to the top hits list.Our third question was to analyze the retweet counts and likes but as discussed with the Professor we realized that retweet count won't be more as we are collecting data of tweets as and when they are created.Hence, that factor was not going to work.

As a future scope, we found out working on a Geo-location of the tweets and analyzing that from where and how many users post a tweet and where the music content is more popular.Due to time constraint, we could not complete that part as such but would like to work on it as future scope.Also, we found out some prediction algorithms with which we could analyze our content and work on predicting new lists.That factor can also be added to our future scope list.

## 7 ACKNOWLEDGEMENT

We would like to Thank our Professor Jeremy H Blackburn for letting us work on this project and guiding us through and helping us with this idea.

## REFERENCES

[1] https://api.twitter.com/2/tweets/sample/stream
[2] https://www.billboard.com/charts/hot-100
[3] https://github.com/guoguo12/billboard-charts
[4] https://developer.twitter.com/en/docs/twitter-api/tweets/sampled-stream/api-reference/get-tweets-sample-stream
[5] https://en.wikipedia.org/wiki/PostgreSQL
[6] https://www.researchgate.net/publication
[7] https://towardsdatascience.com/sentimental-analysis-using-vader-a3415fef7664