Twitter Sentiment Analysis

Nirmal Kanagasabai

***Abstract*—Twitter, one of the most popular micro-blogging service, has been the hot-spot for data scientists to experiment with and find user interests. Through tweets, the users express their opinions on different topics. In addition to those opinions, the users also express their sentiments in this platform. Hence, understanding the sentiment of authors on a given topic has been one of the primary interests of data scientists over the last few years. The purpose of this exercise will be collecting Tweets using Twitter's Streaming API on a Trending topic and score the sentiments of authors on the topic. Three different data corpora (#StrangersThings, #Weather, #USAirlines) were used for the study. To carry out the sentiment analysis, we will make use of SentiWordNet [1] [2], a lexical resource commonly used for opinion mining, and VADER (Valence Aware Dictionary and sEntiment Reasoner) [3], a lexicon and rule-based sentiment analysis tool which is commonly used to analyze sentiments expressed in social media. To evaluate the performance of the two lexicons, a random subset of the data corpora is selected and annotated manually using CrowdFlower. Finally, the accuracy of the sentiment analysis tools on the different corpora and the inferences drawn from the study is discussed.**

I. Introduction

Sentiment Analysis, the process of determining the users' opinions to be positive or negative or neutral, has been carried out for over a decade. This is primarily done to figure out the customer or the follower's attitude towards a product/movie/topic/brand through emotions, contexts, tone, etc. The companies or the stake holders use this study to understand public opinion about them or their products and also measure user satisfaction. The feedback gathered from these studies have been very helpful to influential people and organizations in the past.

Twitter, being the most commonly-used and publicly available micro-blogging platform, has been one of the best areas to tap the users' sentiments. Though several studies have been carried out in the past, this study will review the performances of two simple lexicons (SentiWordNet and VADER) across three different corpora. The following sections throws light on the steps carried out in this exercise: Data Collection, Pre-Processing, Sentiment Analysis using SentiWordNet and VADER, Performance Evaluation, Results and Discussion.

II. Data Collection

Three different data corpora were used in the study. The first dataset was scraped from Twitter using the Streaming API [7] on the topic #StrangersThings2 which was trending on Twitter last week. The second dataset was also scarped from Twitter using Streaming API on the topic #Weather. This task was already done as a part of the sentimental analysis on Weather-Related Tweets. [5]. The third dataset was the Twitter US Airline Sentiment Analysis Dataset that was published in Kaggle [4]. The reason why three datasets were used in this study was to analyze and understand the performance of the Sentiment Analysis tools SentiWordNet and VADER.

Advisor: Dr. Derek Ruths, McGill University, Canada, 2017-2018.

III. Pre-Processing

As the raw-tweets will have so many irrelevant data which might alter the performance of the lexicons, data pre-processing was carried out before scoring the sentiments. The first step was the extract the 'Tweets' alone from the .JSON file. The following steps were carried out during the pre-processing phase:

- Converting all the words to lowercase
- Tokenize the sentences using NLTK Tokenizer
- Remove Twitter Usernames beginning with @ using Regular Expressions
- Removing URLs starting with HTTP or HTTPS using Regular Expressions
- Using Stop Words available in English Language
- Joining Meaningful words after splitting them

IV. Sentiment Analysis

A. SentiWordNet

SentiWordNet [1] is a lexical resource that is predominantly used for opinion mining. To each synset of the WordNet, three sentimental scores are assigned by SentiWordNet: Positivity, Negativity and Objectivity. This pre-calculated scores for the positive, negative and neutral sentiments of some words will help in computing the polarity of the synsets and average the value (for the Tweet).

All steps involved in SentiWordNet's Sentiment Analysis, 'Translation from NLTK tags to WordNet code', 'Part-Of-Speech (POS) tagging of a sentence', 'Word-Sense Disambiguation Technique' which is an adaptation of the technique proposed by Fabio Benedetti [9] is carried out. The main function sentiwordnet_classify(Tweet) breaks the multiple sentences tweets into separate sentences. This helps in improving the context for the word-sense disambiguation technique. A threshold value of 0.75 was defined for Objective words beyond which they are omitted and the polarity is effectively calculated.

B. Valence Aware Dictionary and sEntiment Reasoner

VADER (Valence Aware Dictionary and sEntiment Reasoner) is a lexicon and rule-based sentiment analysis tool that was specifically designed for working on social media texts. They must effectively deliver better performances compared to SentiWordNet. Hence, it was chosen for this study.

In this approach, each of the words in the lexicon is rated as to whether it is positive or negative and in many cases, how positive or negative. Amazon's Mechanical Turk was used by the developers of VADER to rate the words. VADER produces four sentiment metrics from these word ratings nameely: Positive, Neutral, Negative and Compound. The metric, 'Compound' is the sum of all lexicon ratings and is standardized to range between -1 and 1. A truth table was formulated for the three scores (Positive, Negative and Neutral) and the formula to determine the 'sentiment' of the particular tweet was defined.

V. Performance Evaluation

Both the sentiment analysis tools were ran against the the three pre-processed datasets. In order to evaluate the performance of these tools, a comparison had to be made on the predicted sentiments against base / reference sentiment labels.

- In the US Airline dataset, all 14640 Tweets were labeled and hence it was easy to compare the entire lot against the one that's predicted using the SentiWordNet and VADER tools.
- In the Weather dataset, 100 Tweets were labeled already and it was used for evaluation
- In case of StrangersThings dataset, all the Tweets were unlabeled. Hence, a CrowdFlower task was created and 200 Tweets were manually annotated with the appropriate sentiments.

The tables I, II and III (and the Figures 1, 2 and 3) presents the count of Positive, Negative and Neutral sentiments in all the three datasets under study. The sentiment counts were mapped for manually annotated Tweets and the ones predicted by SentiWordNet and VADER. However, this data was not enough to calculate the accuracy of the two sentiment analysis tools.

The predicted Tweets were compared against the manually annotated ones and the count of correctly and wrongly predicted sentiments were presented in Table IV, V and VI. The accuracy was calculated by the formula: (Count of Correct_Sentiments) / (Count of Correct_Sentiments + Count of Wrong_Sentiments) * 100. In other words, the accuracy is the total number of correctly_predicted_sentiments to the total number of sentiments multiplied by 100.

VI. Results and Discussion

The accuracies of the two sentiment analysis tools against the 3 different datasets can be seen at Figure 4. It is to be observed that the accuracy is highest for both the tools (SentiWordNet: 55% and VADER: 66%) in case of Weather Dataset. Likewise, the accuracies have been the least for both the tools (SentiWordNet: 44.77% and VADER: 54.82%) in case of Airline Dataset. It is obvious that the number of manual annotations that was used to compare the performance against the predicted sentiments were the most in Airline dataset (14,640) and the least in Weather Dataset (100). This brings in the question of representation as the chosen sample of 100 Tweets in the Weather dataset may not capture the exact accuracy of the predictions made by these tools. However, not all Tweets can be manually annotated. Hence, a trade-off on the number of samples representing the entire dataset can be chosen to be manually annotated and the performance can be computed against it.

Considering the above mentioned factors, 200 Tweets were chosen in StrangersThings dataset and were manually annotated and the performance comparison gives us optimal results. In both SentiWordNet and VADER tools, the accuracies are more than 50% (53.5% in SentiWordNet and 58% in VADER). VADER, as expected, performs better than SentiWordNet for analyzing Social Media content. It's accuracies on all three datasets have surpassed the SentiWordNet's.

Overall, for the topic 'US Airlines', there were more 'Negative' tweets compared to 'Positive' and 'Neutral' tweets. This could be because of the complaints being raised by the customers on public platforms to form groups of other users who faced similar situations from the air service providers. This study would really help such Air-service providers to look into what were the customers dissatisfied with and address their concerns. Likewise, in case of StrangersThings dataset, there were more 'Positive' tweets compared to the negative and neutral counterparts. It expresses the excitement and the positive sentiments of users towards the TV show.

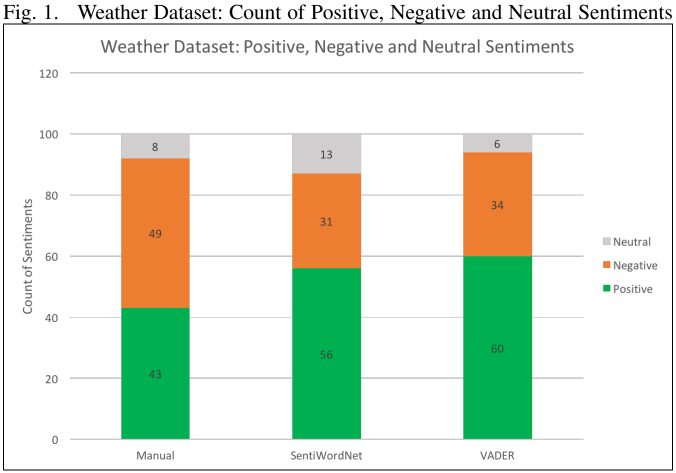
For the Weather dataset, the number of tweets in the positive and negative categories were pretty balanced (close to one another). However, the predictions of SentiWordNet and VADER states otherwise. This brings up an important issue that in such circumstances where you there is an equal balance of positive and negative Tweets, the inaccuracies in the system could drastically change the end results and the users end up making wrong inferences. Hence, It is to be understood that these two tools cannot have accuracies beyond 70% and can be used only as quick and shallow alternatives to an actual Machine Learning classifier.

To identify the keywords that were constantly occurring in most of the Tweets (the terms with maximum frequencies), WordClouds were prepared for all three datasets and for all three labels (Positive, Negative and Neutral). The WordClouds were plotted using Matplotlib and are attached in the Appendix section for the reader's reference (Figures 5 to 13). It can be observed that the Airline's negative cloud has words like, 'Bad Customer service', 'Cancelled', 'Delayed', 'Hold', etc. Likewise, the positive cloud has words like, 'Thank', 'Awesome', etc. The Weather's negative cloud has words like, 'Freezing', 'Crazy', 'Hate', 'Horrible', etc. and the positive cloud has words like, 'Sunny', 'Great', 'Enjoy', etc. One

TABLE I
WEATHER DATASET: COUNT OF POSITIVE, NEGATIVE AND NEUTRAL SENTIMENTS

Polarity	Manual	SentiWordNet	VADER
Positive	43	56	60
Negative	49	31	34
Neutral	8	13	6

Fig. 1. Weather Dataset: Count of Positive, Negative and Neutral Sentiments



important observation in the Word Clouds of StrangersThings would be the inclusion of the hashtag #StrangersThings2 as one of the StopWords to actually identify what the users are talking about. In all three clouds, the hastag has superposed all other keywords.

REFERENCES

[1] Andrea Esuli, Fabrizio Sebastiani. SENTIWORDNET: A Publicly Available Lexical Resource for Opinion Mining *In Proceedings of the 5th Conference on Language Resources and Evaluation (LREC06)*, pp. 417-422, 2006.
[2] Stefano Baccianella, Andrea Esuli and Fabrizio Sebastiani. SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining. *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, ISBN: 2-9517408-6-7, May 2010.
[3] C.J. Hutto. VADER (Valence Aware Dictionary and sEntiment Reasoner). (https://pypi.python.org/pypi/vaderSentiment).
[4] Kaggle: Twitter US Airline Sentiment Analysis Dataset (https://www.kaggle.com/crowdflower/twitter-airline-sentiment)
[5] Nirmal Kanagasabai and Derek Ruths. Analyzing User-Sentiments on Weather-Related Tweets.
[6] Marco Bonzanini *Mastering Social Media Mining with Python*. Packt Publishing. ISBN-13: 978-1783552016.
[7] Twitter Streaming API. (https://dev.twitter.com/streaming/public).
[8] Tweepy - Python Library (http://www.tweepy.org/).
[9] Fabio Benedetti. Tutorial of Sentimental Analysis (https://www.slideshare.net/faigg/tutotial-of-sentiment-analysis).

VII. APPENDIX

TABLE II
AIRLINE DATASET: COUNT OF POSITIVE, NEGATIVE AND NEUTRAL SENTIMENTS

Polarity	Manual	SentiWordNet	VADER
Positive	2363	6659	6196
Negative	9178	5092	5179
Neutral	3099	2889	3265

Fig. 2. Airline Dataset: Count of Positive, Negative and Neutral Sentiments

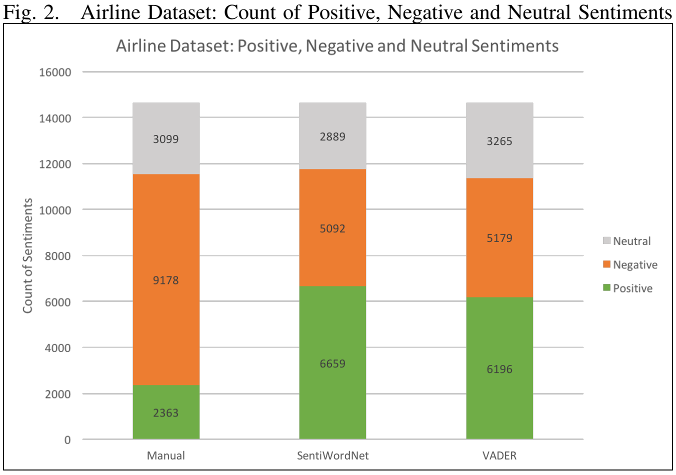


TABLE III
STRANGERSTHINGS DATASET: COUNT OF POSITIVE, NEGATIVE AND NEUTRAL SENTIMENTS

Polarity	Manual	SentiWordNet	VADER
Positive	88	86	71
Negative	64	47	28
Neutral	48	67	101

Fig. 3. StrangersThings Dataset: Count of Positive, Negative and Neutral Sentiments

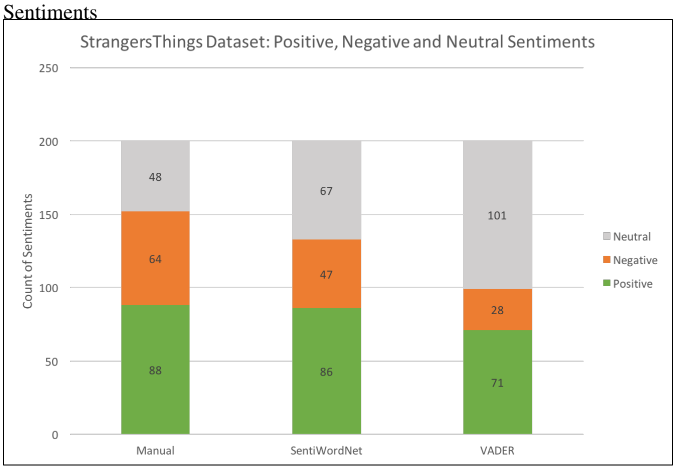


TABLE IV
WEATHER DATASET: ACCURACY

Prediction_Count_Type	SentiWordNet	VADER
Correct Sentiments	55	66
Wrong Sentiments	45	34
Accuracy	55%	66%

TABLE V
AIRLINE DATASET: ACCURACY

Prediction_Count_Type	SentiWordNet	VADER
Correct Sentiments	6555	8026
Wrong Sentiments	8085	6614
Accuracy	44.77%	54.82%

TABLE VI
STRANGERSTHINGS DATASET: ACCURACY

Prediction_Count_Type	SentiWordNet	VADER
Correct Sentiments	107	116
Wrong Sentiments	93	84
Accuracy	53.5%	58%

Fig. 4. Accuracy of SentiWordNet and VADER on Different Datasets

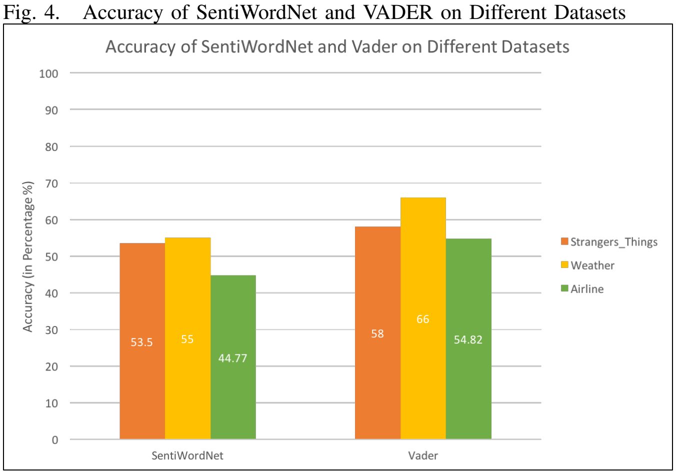


Fig. 5. Airline Dataset: WordCloud of Negative Sentiments

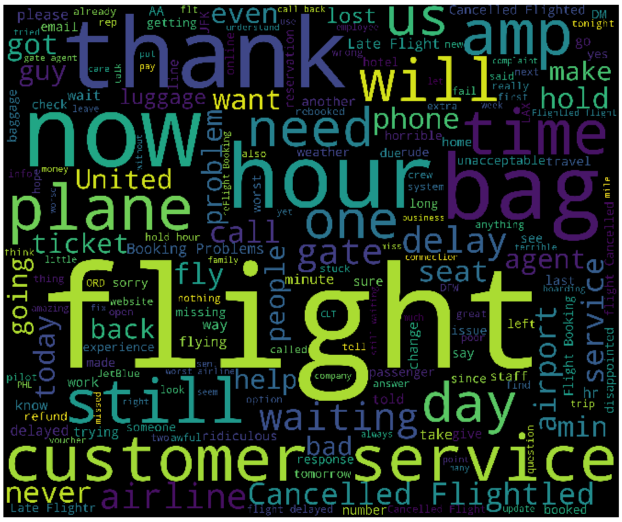


Fig. 6. Airline Dataset: WordCloud of Positive Sentiments

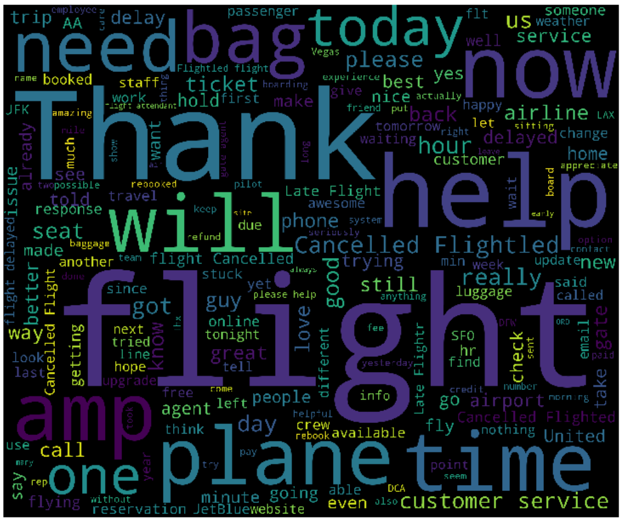


Fig. 7. Airline Dataset: WordCloud of Neutral Sentiments

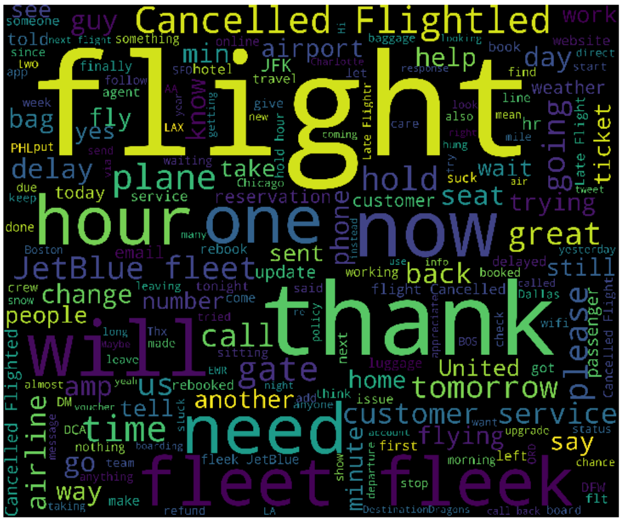


Fig. 8. Weather Dataset: WordCloud of Negative Sentiments

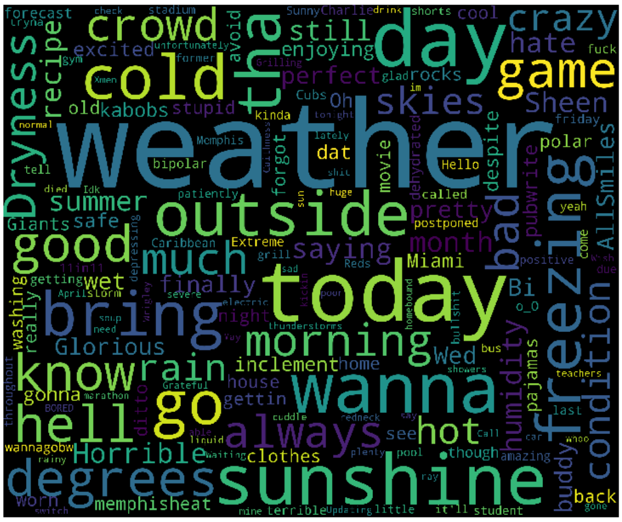


Fig. 9. Weather Dataset: WordCloud of Positive Sentiments

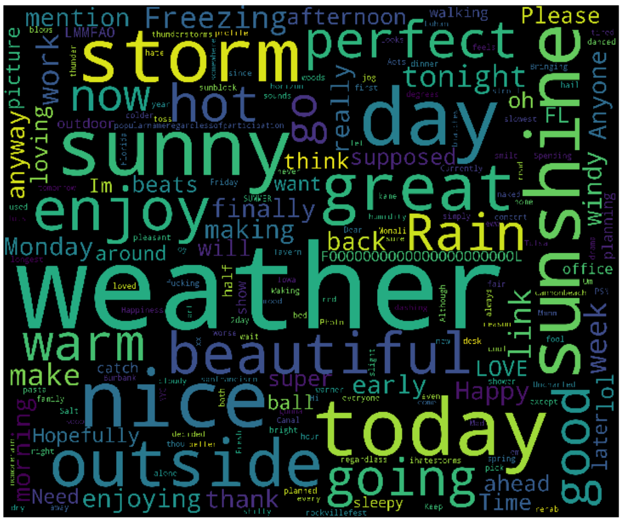


Fig. 10. Weather Dataset: WordCloud of Neutral Sentiments

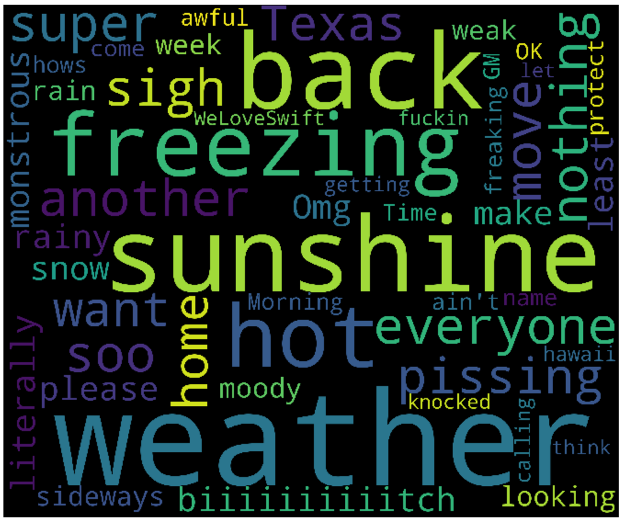


Fig. 11. StrangersThings Dataset: WordCloud of Negative Sentiments

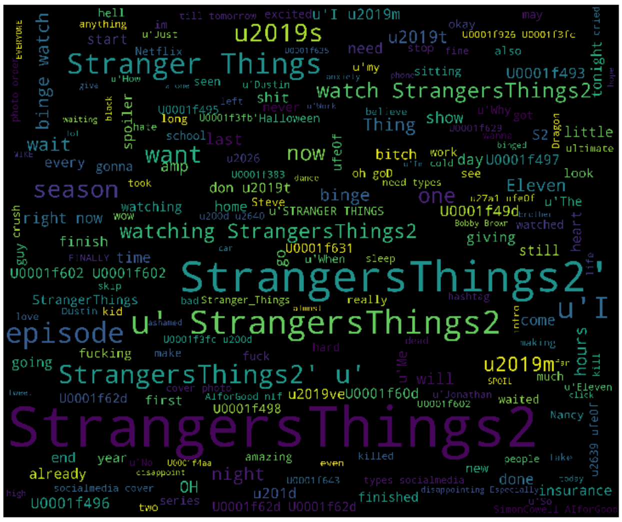


Fig. 12. StrangersThings Dataset: WordCloud of Positive Sentiments

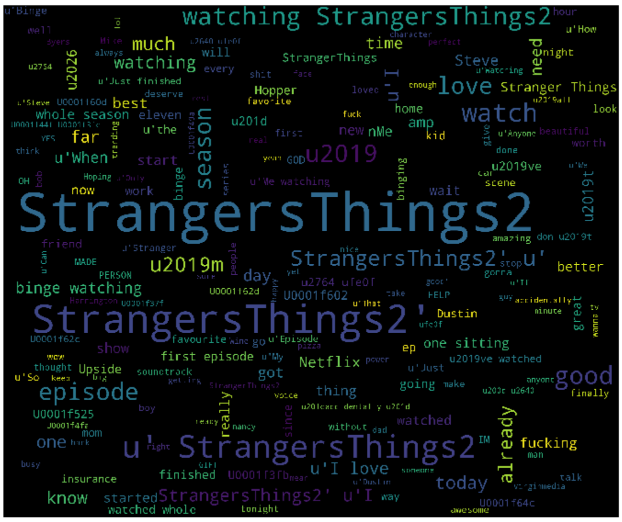


Fig. 13. StrangersThings Dataset: WordCloud of Neutral Sentiments