**Sentiment Extraction**

We employed two approaches to perform the sentiment analysis and assign sentiment scores to our Tesla-related tweets dataset.

The first approach was based on the Naïve Bayes Classifier as described above, which was trained on the Sentiment140 dataset. The second approach was to use The Valence Aware Dictionary and Sentiment Reasoner(VADER) in Natural Language Toolkit(NLTK) using Python[2\_].

**Naïve Bayes Classifier**

We applied the Naive Bayes classifier trained as above to the Tesla-related tweets dataset. The Naive Bayes model classified tweets into two categories: positive(1) and negative(0). Figure X shows a snippet of the sentiment classification results and the allocation of the tweets of the two polar sentiments. We then calculated the percentage of negative tweets for each day. This information can be used as a predictor for the Tesla stock return prediction modeling in the next section.

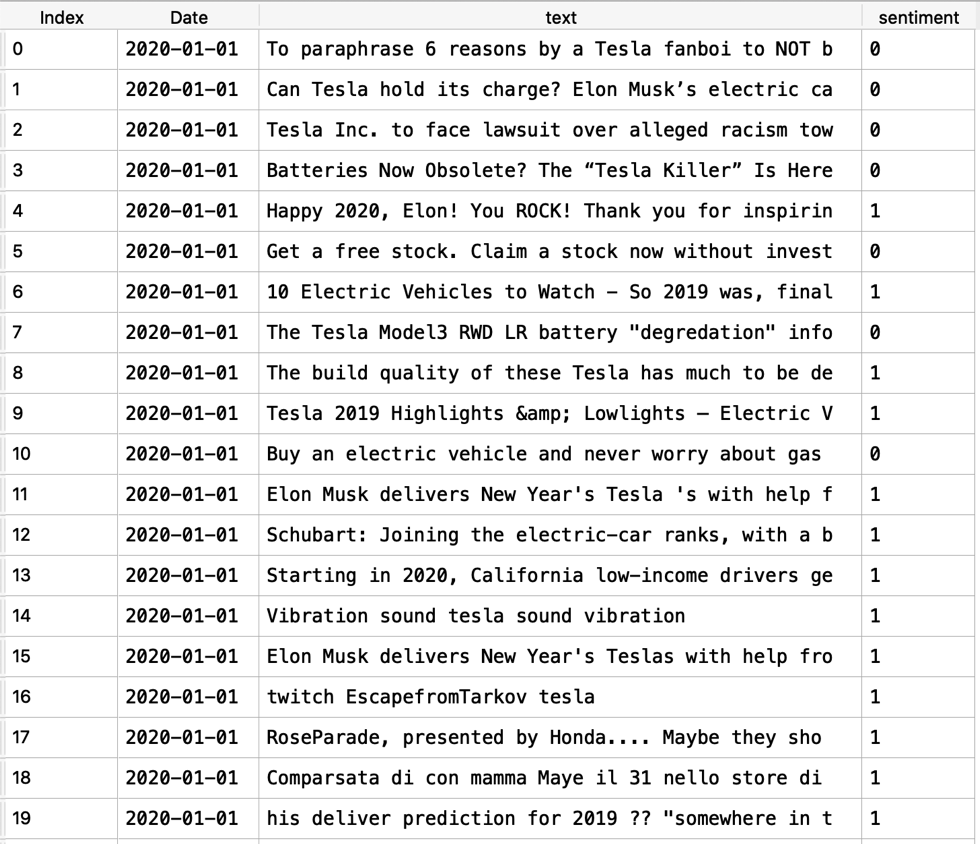
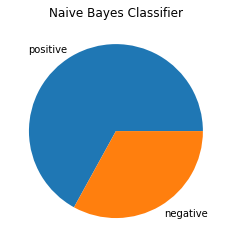
 

Figure X. Polar Sentiment Results using Naïve Bayes Classifier

**VADER**

VADER, unlike the first approach, is a lexicon-based approach, which does not need to train. The VADER sentiment scores are calculated based on its pre-prepared lexicon of sentiment-related words. Using its specially designed VADER model, we can calculate the sentiment scores for the concatenated tweets of each day. Worth mentioning that VADER is a great choice since it is specially attuned to social media contexts. It takes into consideration of emoticons, slang, intensifiers, acronyms, etc., in addition to commonly used sentiment-related words. It has proven results in the use of tweets dataset as shown in the original VADER paper. [1\_]

After understanding VADER, we applied VADER to the Tesla-related tweets dataset and got the sentiment scores for the tweets of each day. We used the three scores from the VADER outputs: positive, negative, and neutral scores. We did not use other VADER output elements such as the compound score since many tweets had the same compound score, which would not be helpful in the modeling process later.

Figure XX shows the polar sentiment scores using VADER, and the allocation of sentiments in the Tesla tweets.

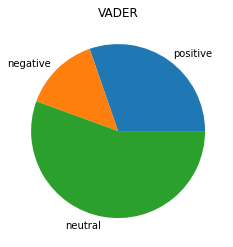
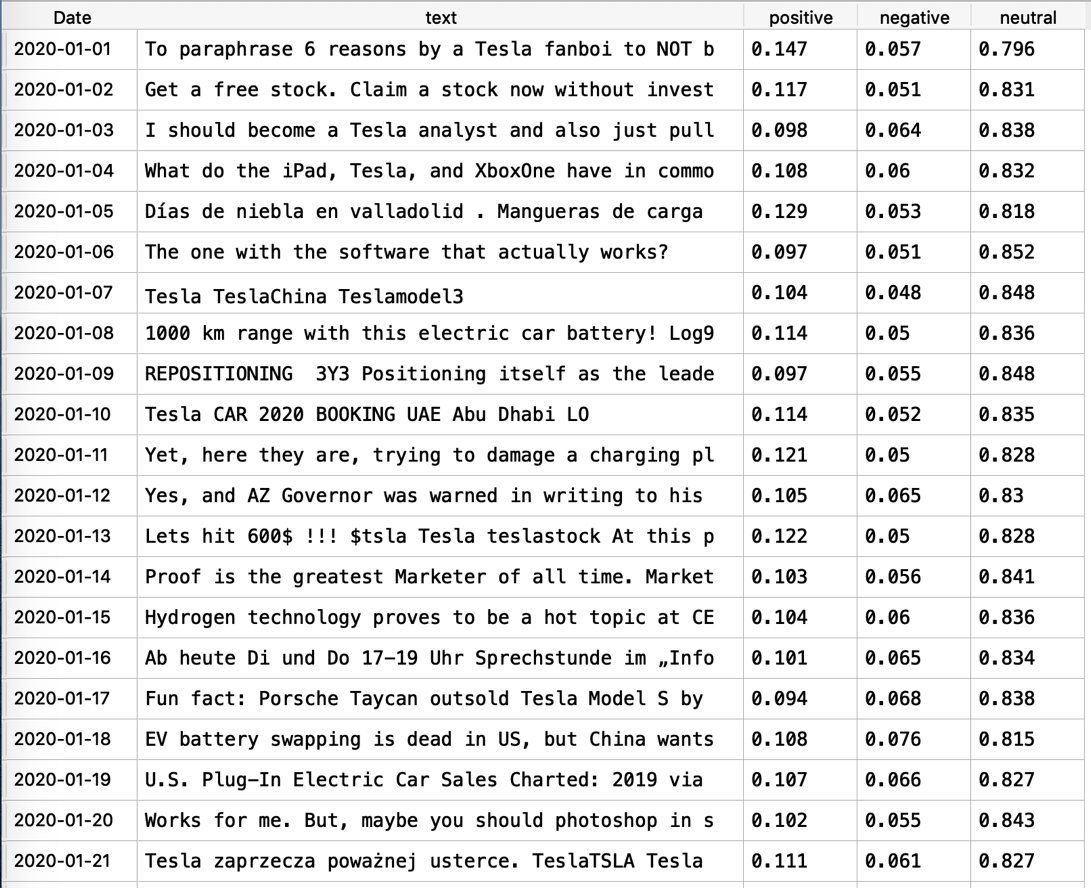


Figure XX. Polar Sentiment Results using VADER

Ref:

[1\_] VADER Paper

<https://www.aaai.org/ocs/index.php/ICWSM/ICWSM14/paper/view/8109/8122>

[2\_] VADER in NLTK

<https://www.nltk.org/index.html>