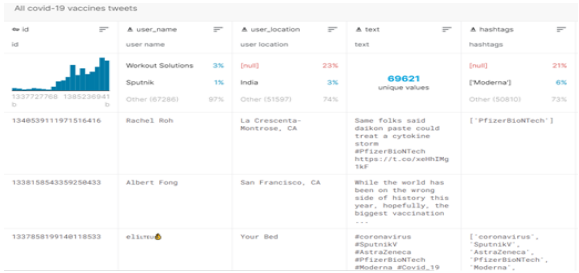
sentimental Analysis of COVID-19 Vaccines Based on Twitter

Team Members: Yifu Qiao, Fengnian Guo, Xiaohan Liu, Daniel Scott

1. Introduction & Dataset

- Goal: Explore different attitude towards different types of COVID-19 vaccines according to Twitter.
- Dataset: Tweets about Vaccines:



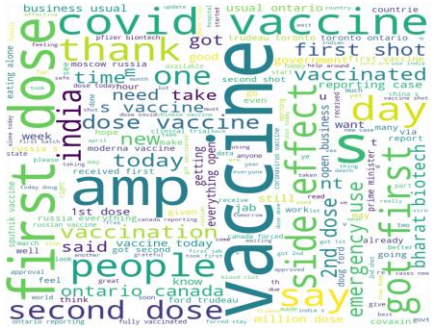
- Dataset for training: Tweet Sentiment Dataset
  Vaccine dataset does not have sentiment labeled, we need to use another dataset to train our classifier.

2. Preprocessing

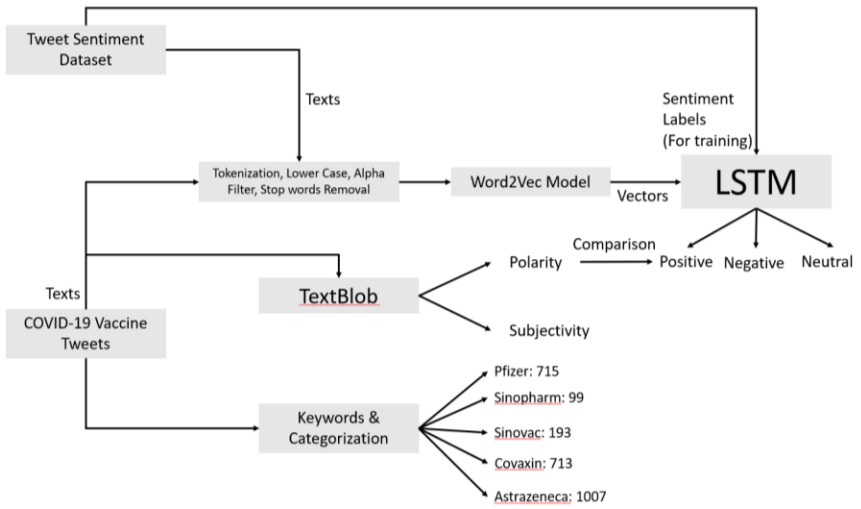
Eliminating tweet tokens: tweet-preprocessor
Tokenization: word tokenization, lowercase, stopwords, alpha filter

3. Categories & Keywords

- Pfizer(715) - Sinopharm(99) - Sinovac(193)
- Covaxin(713) - Astrazeneca(1007)
- Keyword extraction: Top Frequency & TF-IDF



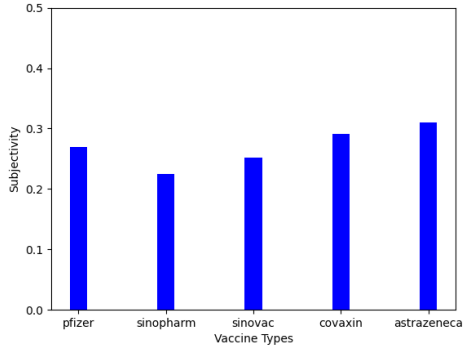
Flow Diagram



4. Sentiment Analysis (TextBlob)

By using the module of TextBlob, we can know the polarity and subjectivity of a review:
- Subjectivity: float number from 0 (subjective) to 1 (objective).

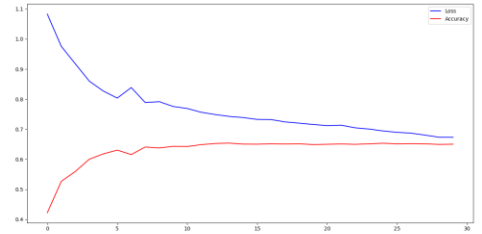
We can compare the polarity results to the classifier results.



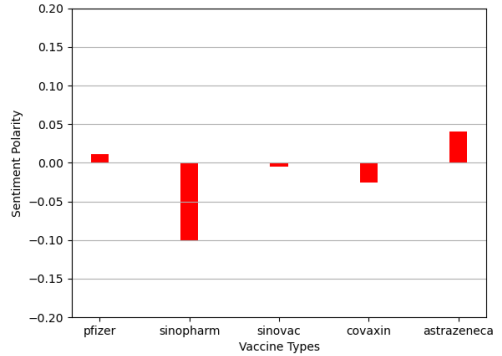
5. Sentiment Analysis (Classification)

- Use Word2Vec to vectorized preprocessed tokens. (Skip-gram algorithm)
- Vectorized Tweets are inputs for LSTM model.

Learning Curves



- Trained LSTM outputs sentiment labels (3 Types). Then we can get sentiment polarities for them:



Statistics of Performance

	Precision	Recall	F-measure
Positive	0.743	0.680	0.711
Negative	0.657	0.543	0.595
Neutral	0.586	0.697	0.637

References
1. https://pypi.org/project/tweet-preprocessor/
2. https://www.kaggle.com/gpreda/all-covid19-vaccines-tweets
3. https://www.kaggle.com/maxjon/complete-tweet-sentiment-extraction-data
4. https://textblob.readthedocs.io/en/dev/ 5. https://en.wikipedia.org/wiki/Word2vec 6. https://pypi.org/project/nltk/ 7. https://pytorch.org/