Better Understand your Consumers via Text Emotional Sentiment Analysis

By Lucas Fishbein

Understanding the Problem

Social Media is the new Town Square

People more and more use social media to express themselves

A company can receive 1000s of tweets

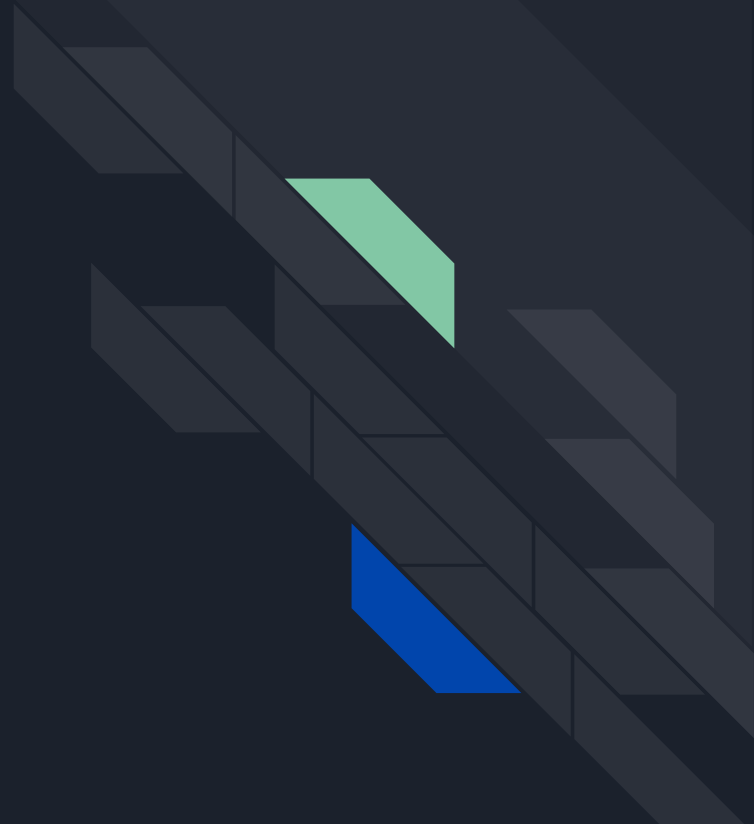Through COVID-19 Vaccinations have become a political Issue

Use the Info consumers are providing

Making use of the info locked in a flood of tweets is difficult

First step is to Organize tweets, then tweets can be scraped of info for analysis

Project Goal:
Use Tweets to Gain a Greater Understanding of Public Opinion and Consumerbase

Project Overview and Goals

- Build a Model to classify Tweets via emotional sentiment

- Organize tweets in order to utilize information

- Identify Groups that already have positive opinions towards a product/brand

- Display the types of deeper analyses that can be run once this data is classified

Data Source Understanding

○ Present dataset contains over 9000 tweets, pre-labeled by human raters for tweet sentiment and sentiment focus

○ Tweets were collected after and focused on SXSW 2011
  a. Sourced from CrowdFlower via [data.world] added by Kent Cavender-Bares on August 30, 2013.
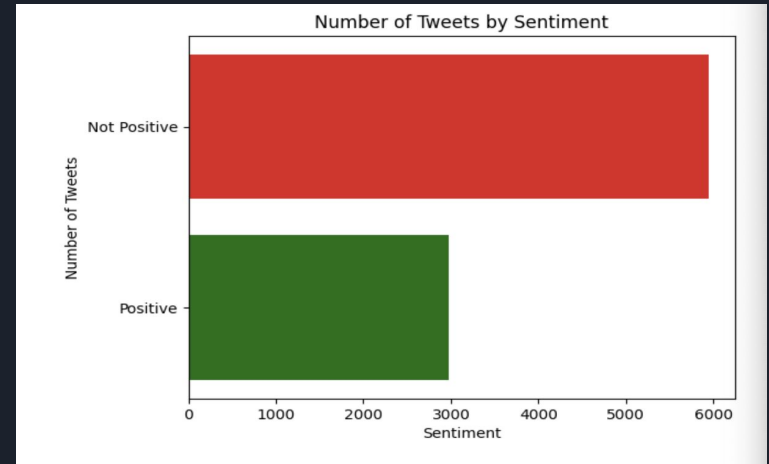
Tweet Text Preprocessing for Classification

- The only input required is Raw Tweet Text directly from twitter,
  - The text is then cleaned of extraneous characters
  - Reformatted
  - Tokenized
  - Lemmitized
  - Converted into padded arrays and input into the model.

- The resulting output is a tweet labeled with its sentiment as "Positive" or "Not Positive"
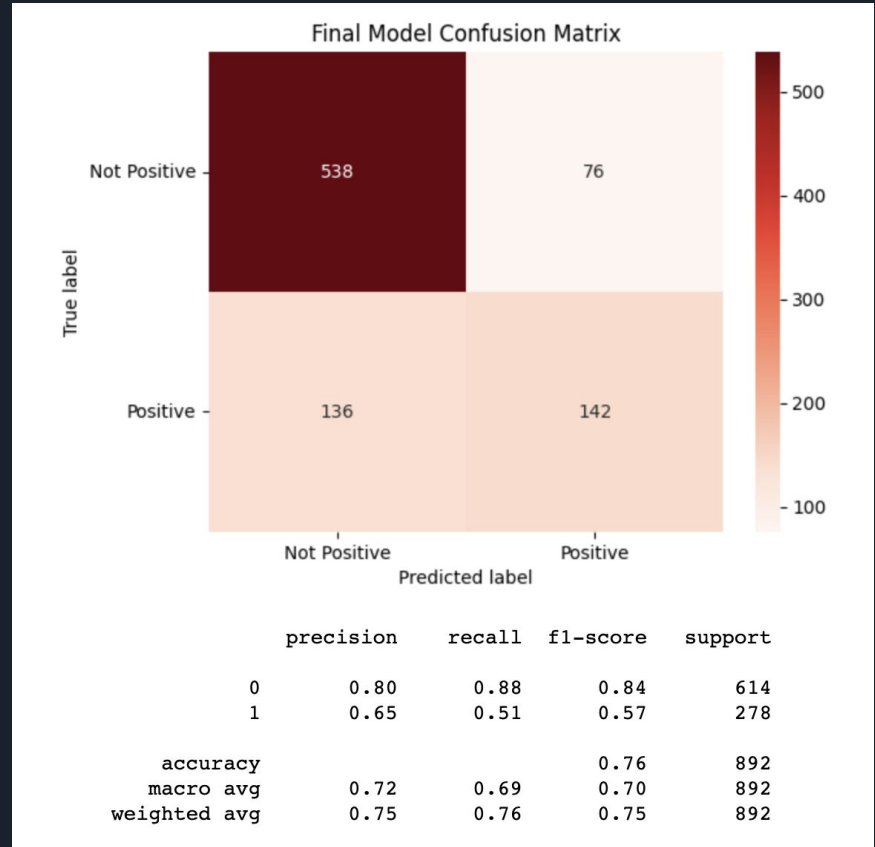
Building a Binary Classification Model

- Two different CNN NLP models were built out, trained, tuned and compared
  - The Best performing epoch from the best performing model became final model

- Model's Performance
  - 76% Overall Predictive Accuracy
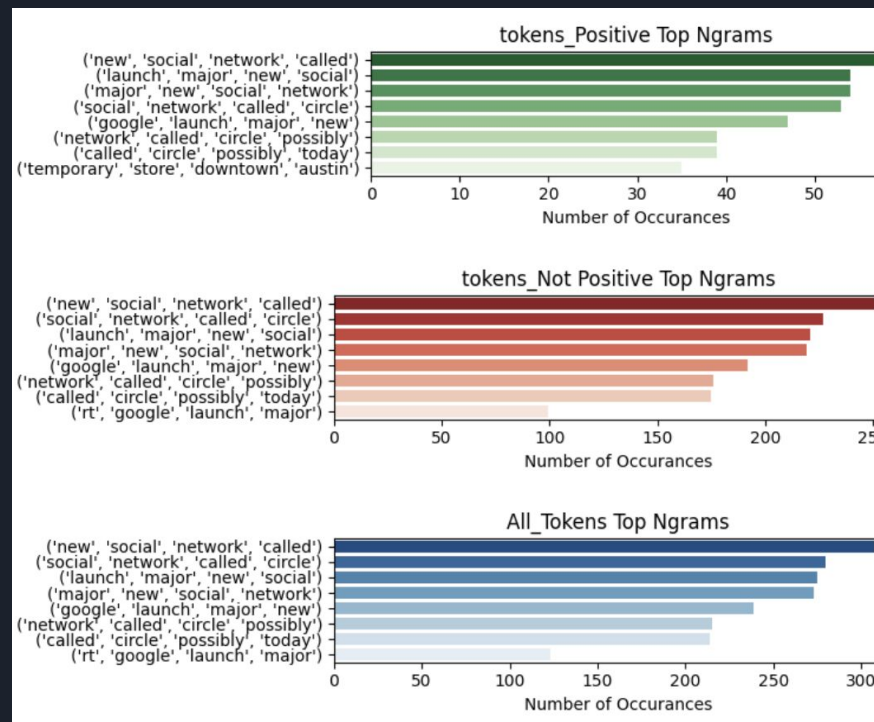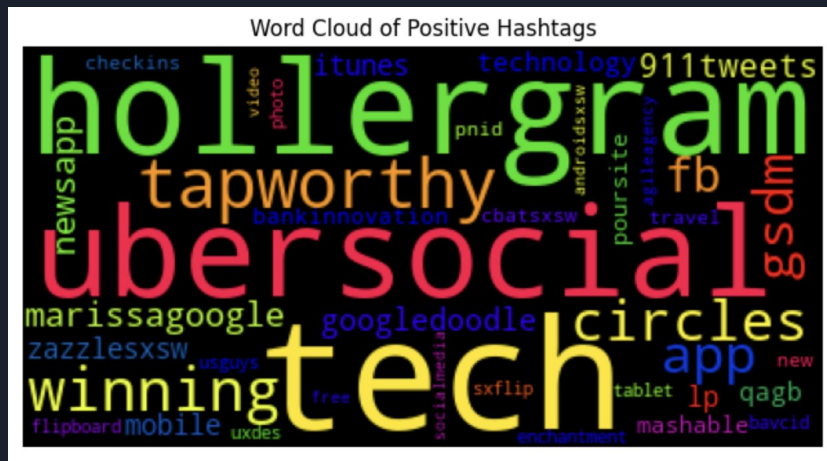 - Dataset was partitioned prior to model training and performance was evaluated via a dataset novel to the model

Final Model Evaluation

- Model has an overall predictive accuracy of 76%
- 88% correct on "Not Positive"
- 51% correct on "Positive"
- Model is significantly better at classifying "Not Positive" tweets

Final Model Confusion Matrix

	Not Positive	Positive
Not Positive	538	76
Positive	136	142

	precision	recall	f1-score	support
0	0.80	0.88	0.84	614
1	0.65	0.51	0.57	278
accuracy			0.76	892
macro avg	0.72	0.69	0.70	892
weighted avg	0.75	0.76	0.75	892

Extracting Info from Labeled Data Examples

1. **N- Gram Analysis for common Phrases**

2. Hashtag Exploration

3. Brand Tweet is directed at

Word Cloud of Positive Hashtags

tokens_Positive Top Ngrams

tokens_Not Positive Top Ngrams

All_Tokens Top Ngrams

Recommendations

From Hashtag Analysis:

- The "Hollergram" campaign received a lot of positive tweets
  - Continue this campaign

- The "Infektd" product received a lot of not positive tweets
  - Rethink the product/ marketing

From N-Gram analysis:

- "Apple is opening temporary store" was a top phrase within not positive tweets
  - Further research should be done to see why this is the case

- Multiple top positive N-Grams contained the phrase "new technology"
  - New releases can likely be used to boost positive opinion

Examples of Practical Uses of Model

General Public Sentiment/ Reputation Management

- Understand how the public generally feels about a brand/product

Utilize Customer Feedback

- Understanding Customer feedback helps identify strengths and areas of improvement

Gain Understanding of Customers

- Understand consumer behavior and develop targeted ad campaigns
-

Easily Retrain Model for other Uses

- Inputting a different training dataset will allow for more targeted analysis if required

Contact Info and Github Repo

Lucas Fishbein of the Flatiron School Data Science Program

Located at 248 Address Street, New York NY 10007

Github Repository: https://github.com/LucasFishbein/Positive_Tweet_Sentiment_CNN_Model/

FishbeinLucas@gmail.com

(129) 921-1234