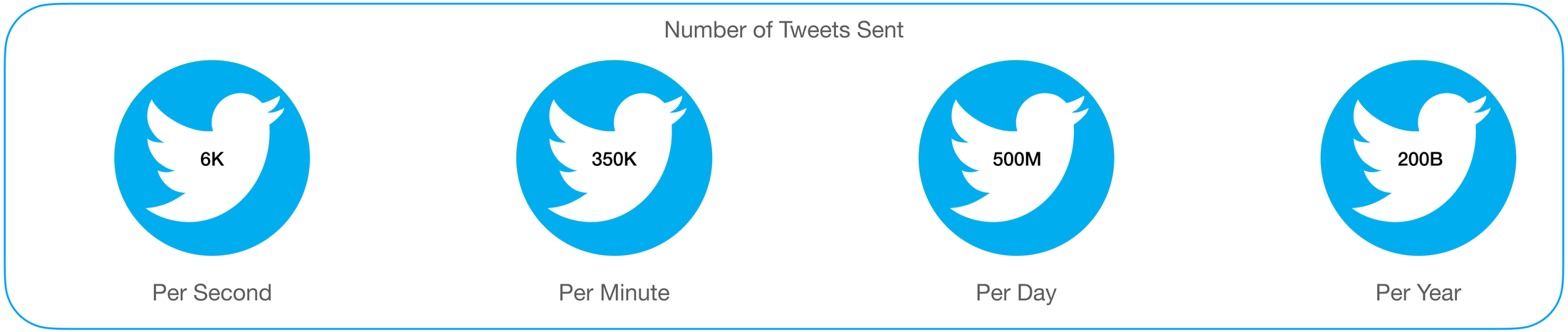
Twitter Sentiment Analysis

Context / Project Overview



- Data available through twitter is not only public and large, but also forever growing
- Analyzing this data can help provide real-time views into consumer sentiment and feelings
- Classification provides product managers / public relations professionals a way to quantify qualitative tweet data

https://www.internetlivestats.com/twitter-statistics/#trend

The Data

Overview of Tweet Data

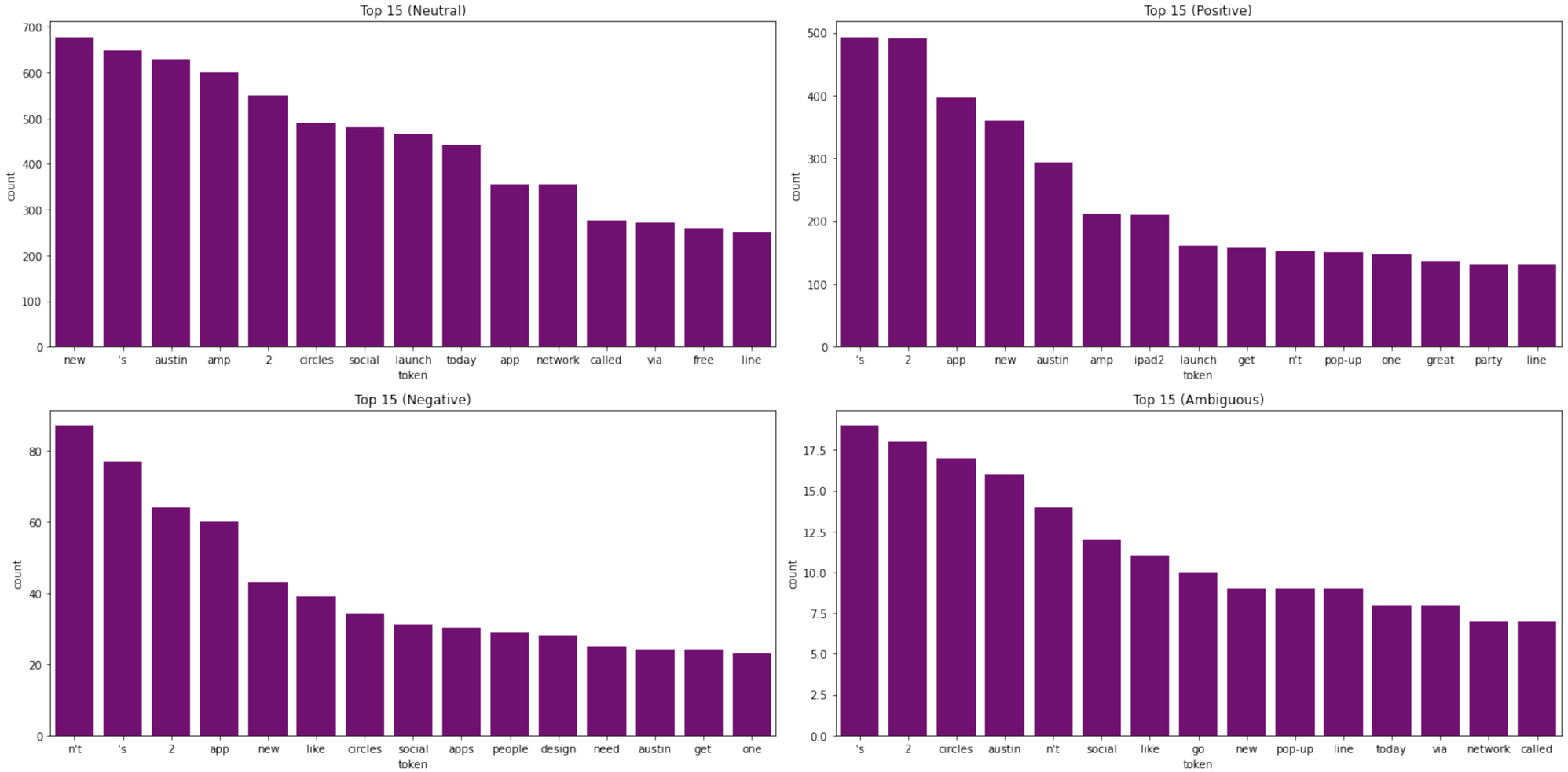
.@wesley83 I have a 3G iPhone. After 3 hrs tweeting at #RISE_Austin, it was dead! I need to upgrade. Plugin stations at #SXSW.	*Negative emotion*
@jessedee Know about @fludapp ? Awesome iPad/iPhone app that you'll likely appreciate for it s design. Also, they're giving free Ts at #SXSW	*Positive emotion*
@swonderlin Can not wait for #iPad 2 also. They should sale them down at #SXSW.	*Positive emotion*
@sxsw I hope this year's festival isn't as trashy as this year's iPhone app. #sxsw	*Negative emotion*

Average Tweet Length: 24.4 Tokens | Total Vocal Size: 13,212 Unique Tokens

9,092 Non-Null Text Entries	5,389 Neutral Labels	2,978 Positive Labels	570 Negative Labels

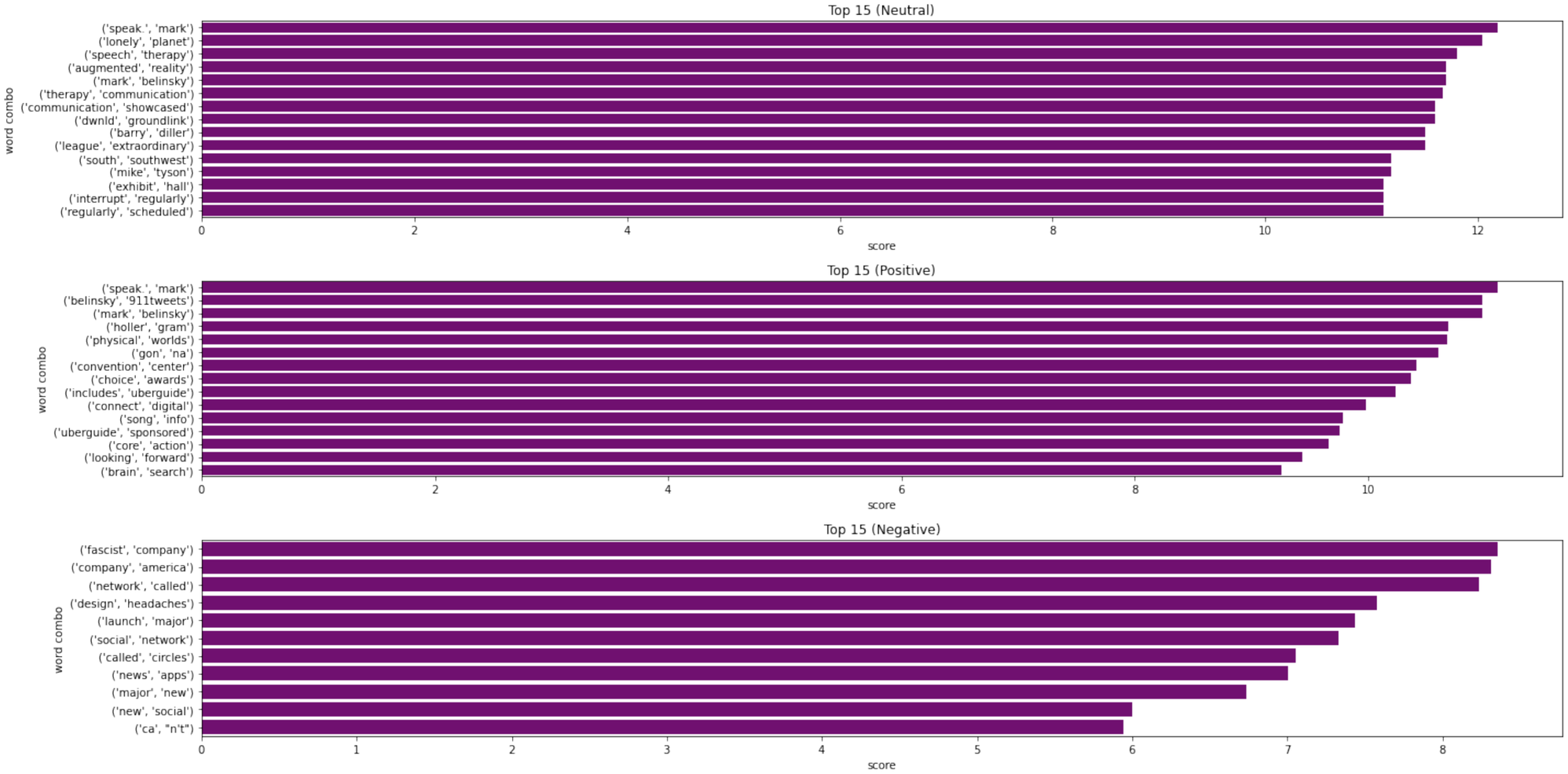
The Data

Most Common Tokens by Class Label



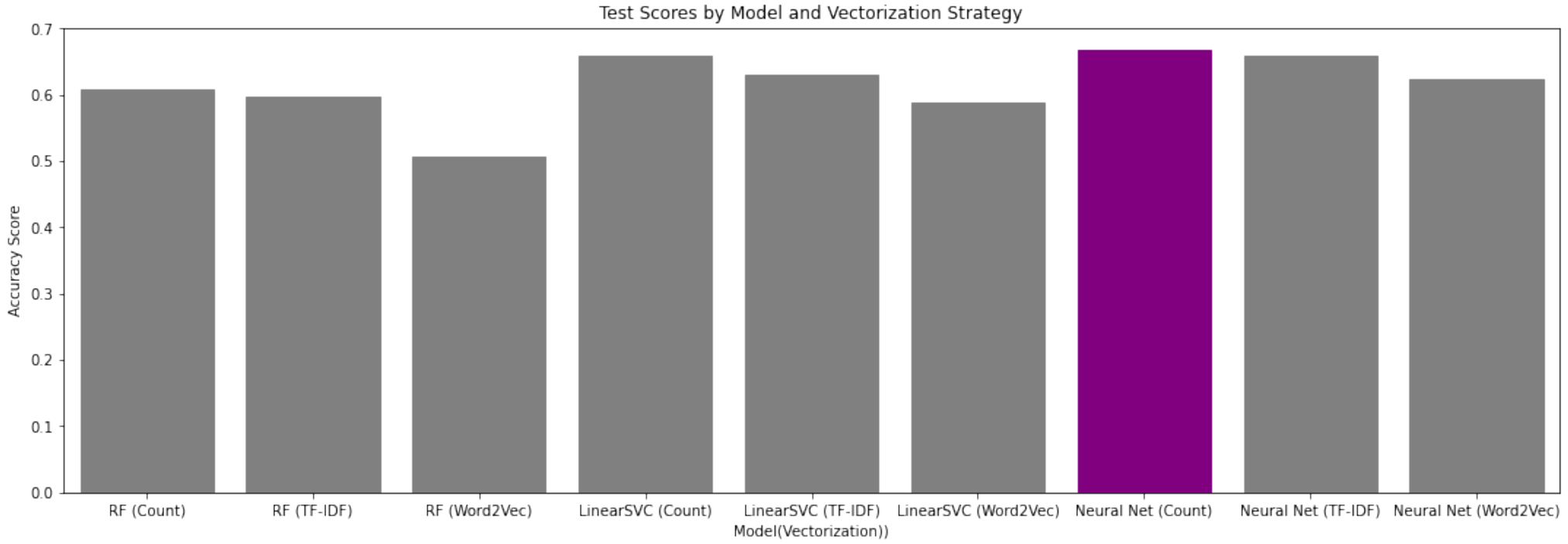
The Data

PMI Scores by Class Label



Evaluation

Model Performance



Final Train Accuracy: 77.7% | Final Test Accuracy: 67.2%

Recommendations

- Run batch of tweets through identified classifier in conjunction with major product releases
- Use classifier to count the number of positive vs. negative tweets related to the company / product that are being released every x time period
- Classifier provides product managers and public relations professionals to track sentiment outside of traditional streams (reviews, surveys, online ratings, etc.)

Areas for Further Work

- Build API to pull tweets directly from Twitter
- Build out dataset to include more positive and negative class labels
- Layer classifier into broader data science process - potential ideas include: keeping a real-time running consumer sentiment score, layer tweet classification into broader time-series model and relate back to stock price, use classifier and build other analytics and tie to company performance
- Combine with some customer service functions and data scraped from completed surveys / consumer reviews

Thanks