Tweet Sentiment Analysis

Alejandro Colocho

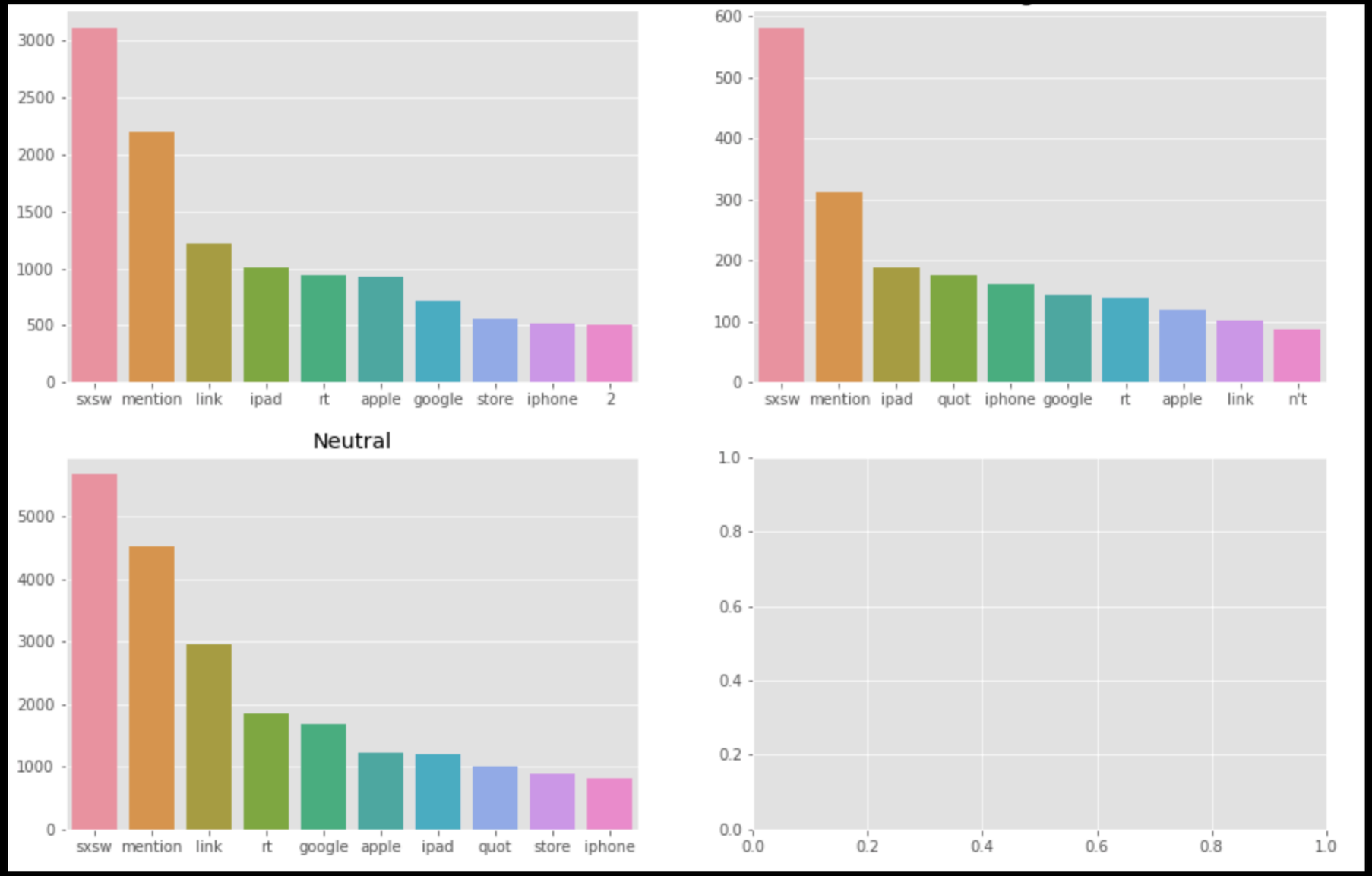
Data Description

- 9,032 tweets regarding Apple and Google.
- Include:
  - Tweet text.
  - Sentiment.
  - What the sentiment is directed towards.

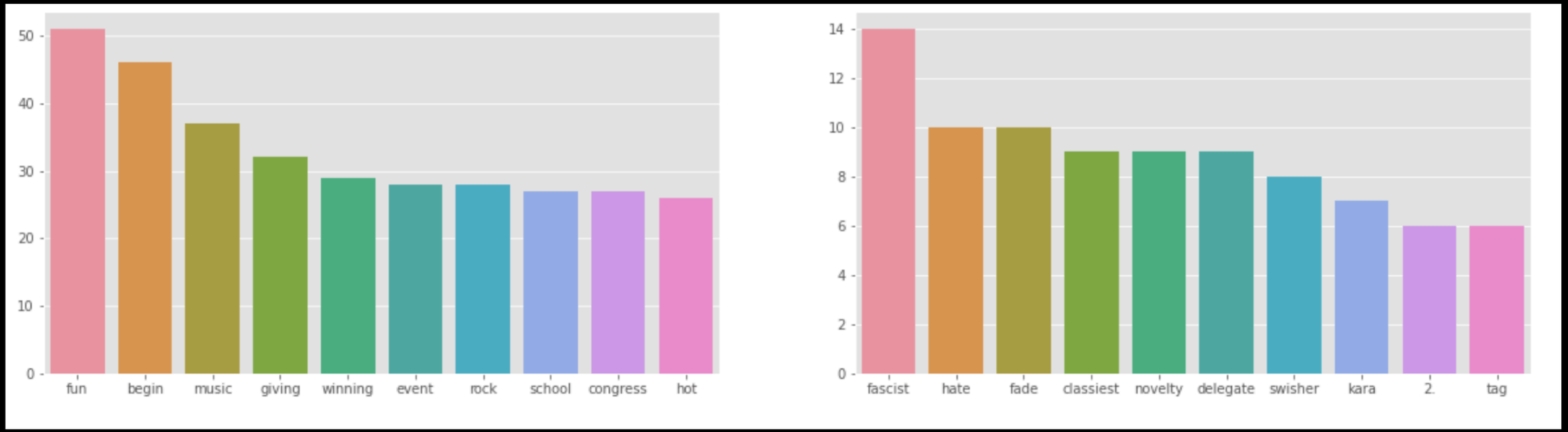
Questions to explore.

- Can we successfully classify positive and negative tweets?
- Can we successfully classify positive, negative, and neutral tweets?
- What are the most import words for each category?

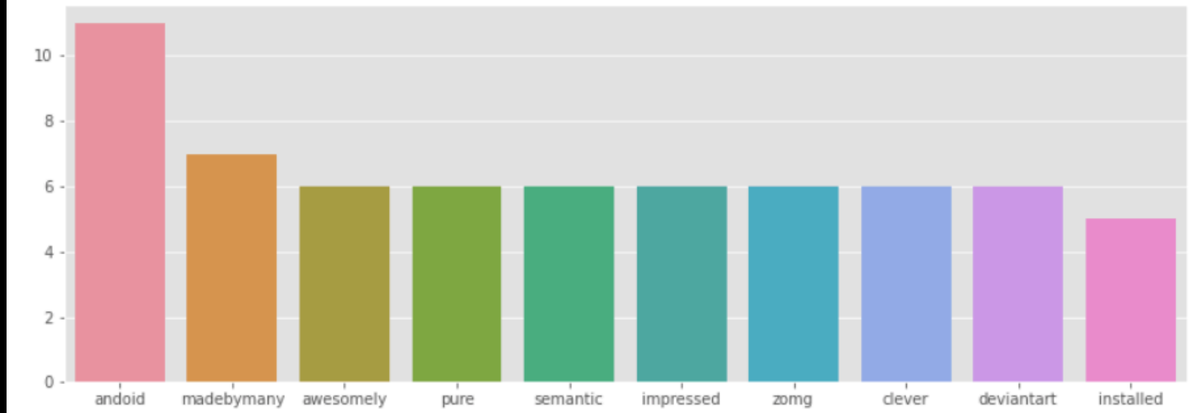
Most Used Words

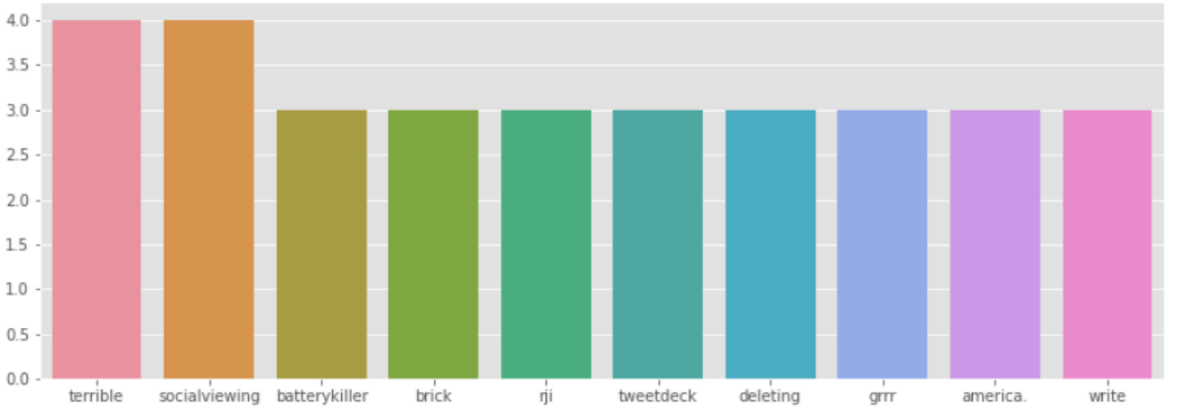


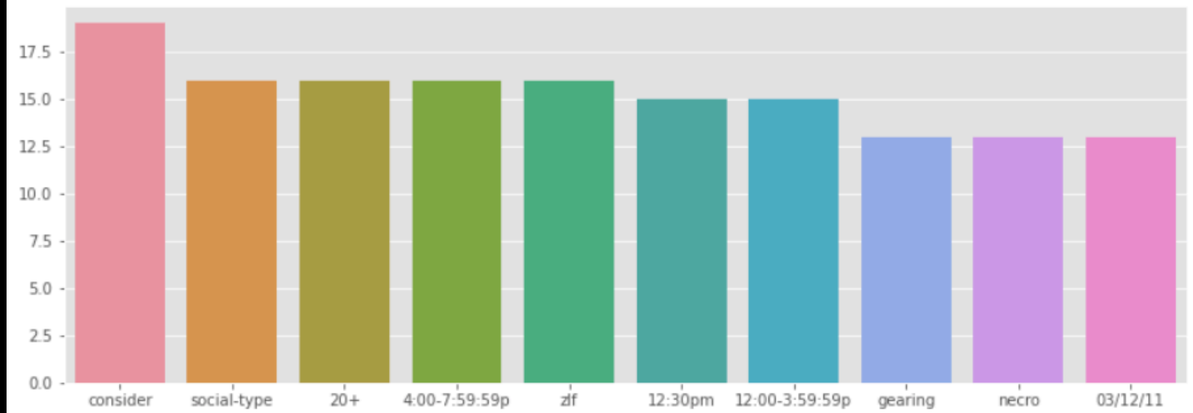
Positive and Negative Important Words

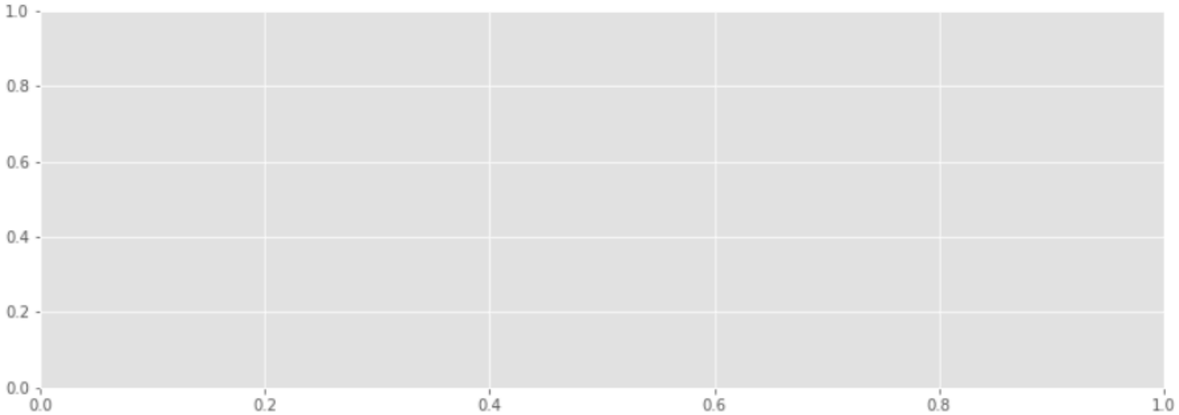


All Important Words





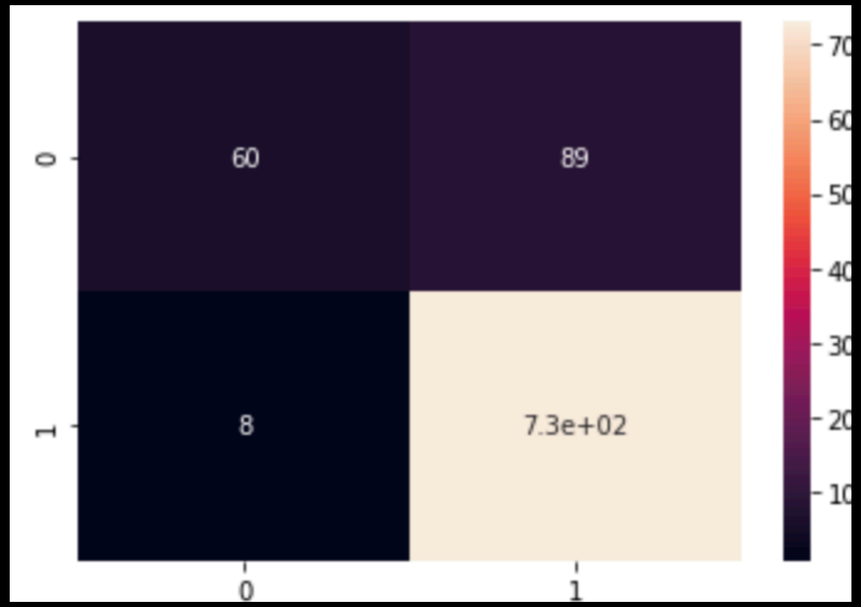




Positive-Negative model.

- Tuned Random Forest model was able to achieve an accuracy of 89%
- The data used was data that eliminated common words.
- It is believed that accuracy was so high because the model was biased towards the positive category. More data is needed to improve the model.

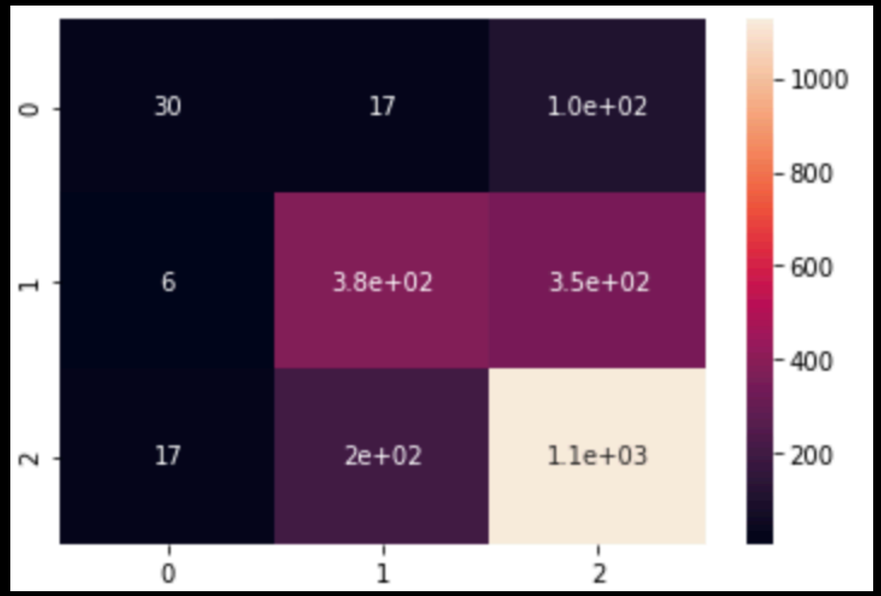
Iteration	1	2	3	4
Accuracy	.87	.88	.88	.89
Model	SVC	Random Forest	SVC (All Data)	Tuned Random Forest



Positive-Negative-Neutral Model

- Tuned SVC achieved the same accuracy as the untuned version. However, it slightly classify more tweets correctly.
- The model performed better when all data was used.
- The model was biased towards neutral tweets. More data is needed to improve the model.

Iteration	1	2	3	4
Accuracy	.69	.67	.69	.69
Model	SVC	Random Forest	SVC (All Data)	Tuned SVC



Limitations of Data

- Imbalanced classes.
- Sentiment directed towards data is mostly incomplete.

Future Work

- More data.
- Tweets directed towards similar companies.
- Look at likeliness of each tweet pertaining to each category.

Conclusion

- Although the positive vs negative model performed really well, it is likely that the model was biased, and it cannot reliably predict never before seen data.
- More data will help to make both models better, so more collection needs to be done.
- Positive and negative tweets have important features that can aid future development and help understand where a tweet falls in a positive-negative spectrum.
- Neutral tweets are not reliable.

Questions?