Pay Attention to this Bag of Tweets!

Eric Adsetts and Eryk Wdowiak

Context

- How can brands analyze the way people feel about them by analyzing mentions on social media?
- We can use NLP to classify tweets by sentiment (Positive, Negative, Neutral)
- Used tweets about Apple and Google from the SXSW festival
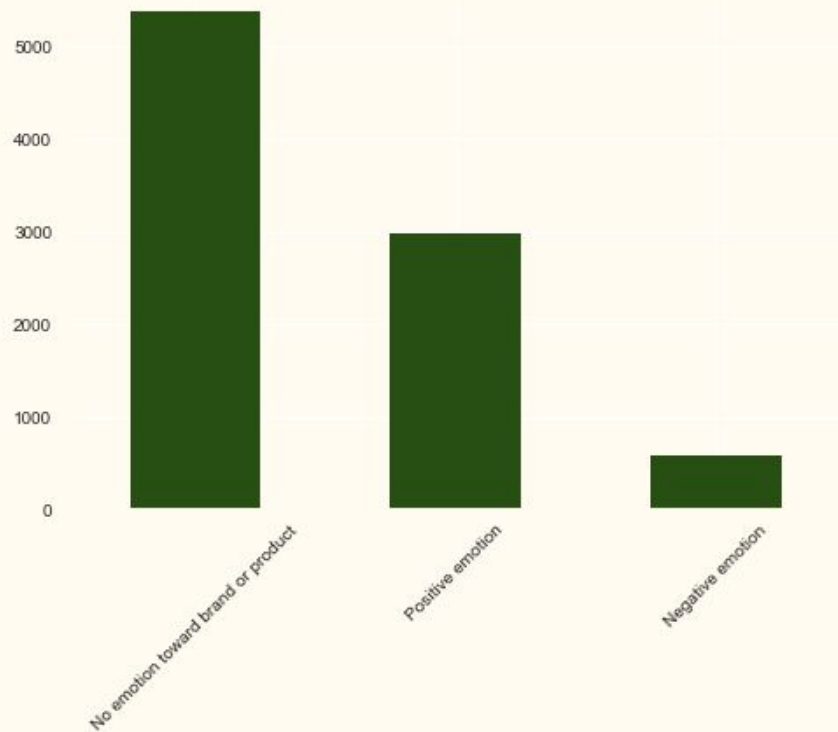- Able to classify tweets with an f1 score of .7
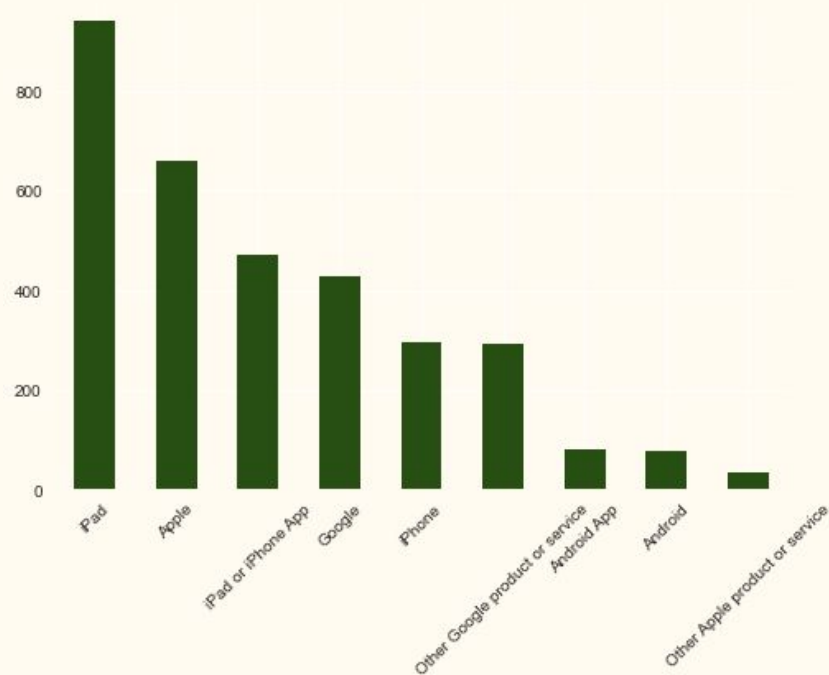
Our Data

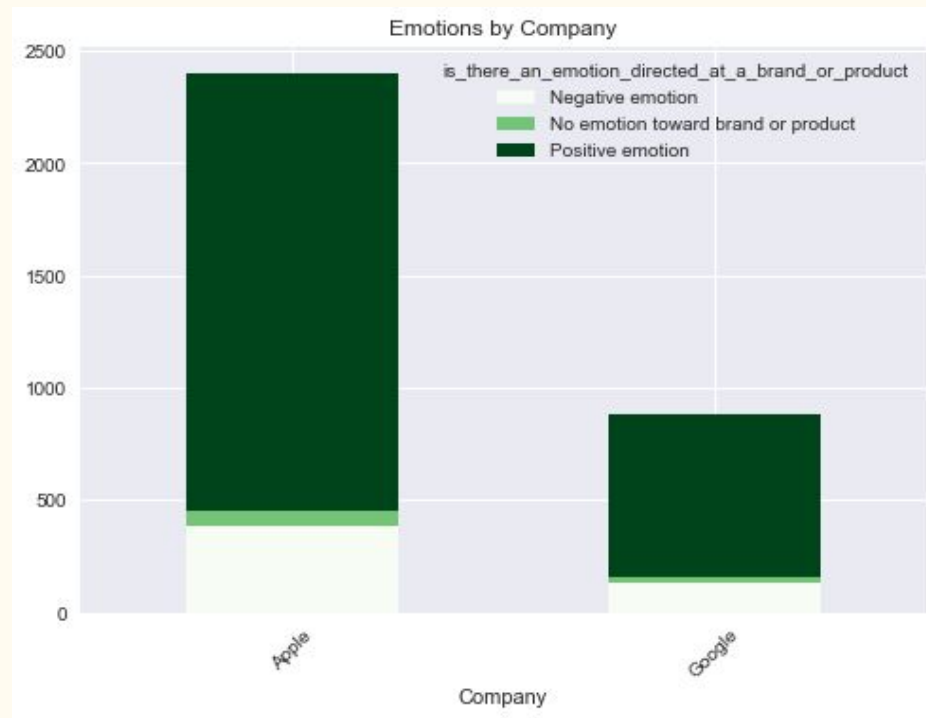9092 Tweets
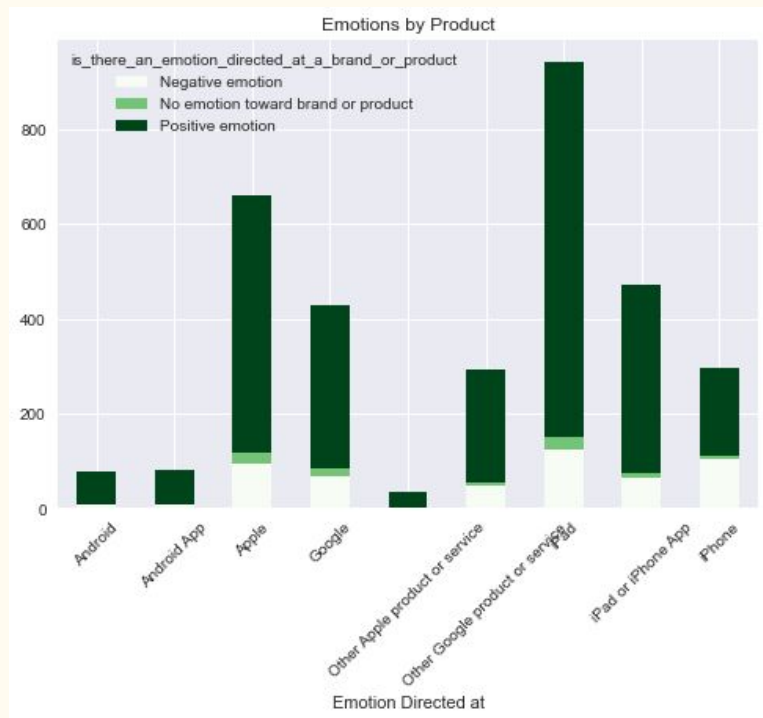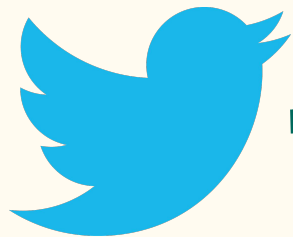
Mostly Neutral

Mostly about Apple

Our Data

Our Data

Process

Data from tweets about Apple and Google

Dating Cleaning and EDA

Lemming, vectorization and n-grams

NLTK

Visualization with Matplotlib and Seaborn

Modeling

Results

Models	F1 Score
Naive Bayes	0.618
Random Forest	0.643
Logistic Regression	0.667
Decision Tree	0.598
K-Nearest Neighbors	0.609
Support Vector Machine	0.667
XG Boost	0.651
Voting Classifier	0.653

SVM Heatmap
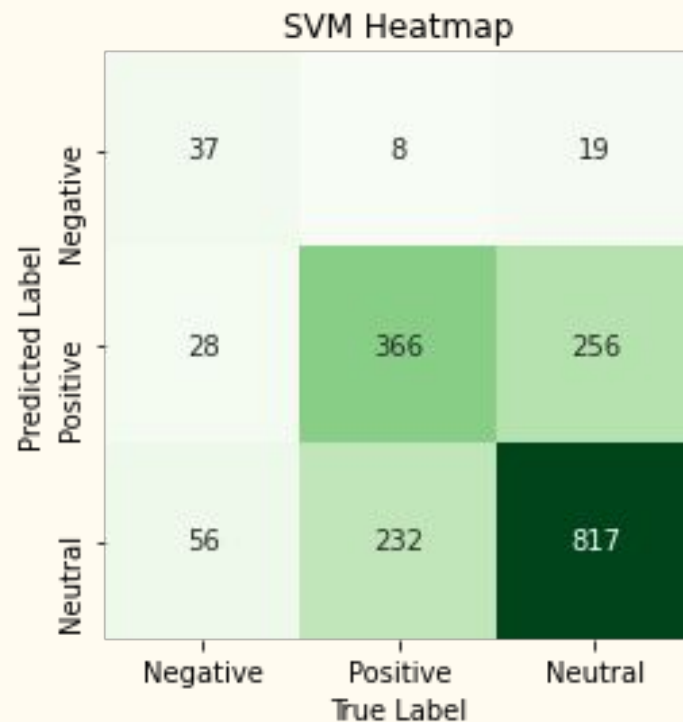
Self-Attentional Sentence Embeddings (1)

- Lin, Feng, et al. (2017) combine word embeddings and self-attention
- Word Embeddings
  - represent a word's semantic information in multidimensional space
  - semantic information can be used to classify sentiment
- Self-Attention
  - self-attention layers model the relationships between words in a sentence
  - those relationships can be used to identify the most important words

	this	phone	is	amazing
att0	0.84	0.16	0.0031	1.4e-05
att1	9.2e-05	0.00032	0.0047	0.99
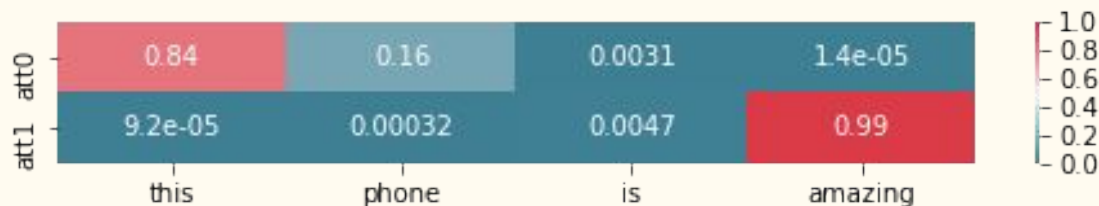
Self-Attentional Sentence Embeddings (2)

- If relationships among words were stationary, we could use this method to:
  - Classify
  - Understand that classification
- The trouble is that relationships among words is not stationary.
- So we will need more time to implement this approach.

Applications

- A company can use this information to measure sentiment towards their products and their competitors
- Could measure success of marketing campaign based on twitter sentiment

- Next steps: further explore sentence embeddings to improve our model
- But for now, "bag of words" methods work better, so:

Pay Attention to this Bag of Tweets!