



# Tweet NLP



Asher and Austin



# Overview

---

Build a model that can rate the sentiment of a Tweet based on its content.

This can be used to show approval ratings for new products in the tech world and other events.

The data is unbalanced with a heavy skew in the neutral emotion.

# Data

The dataset comes from CrowdFlower via [data.world](#)

Human raters rated the sentiment in over 9,000 Tweets as positive, negative, or neither.

Using this data we were able to break down the word usage and predict the emotion perceived from each tweet.

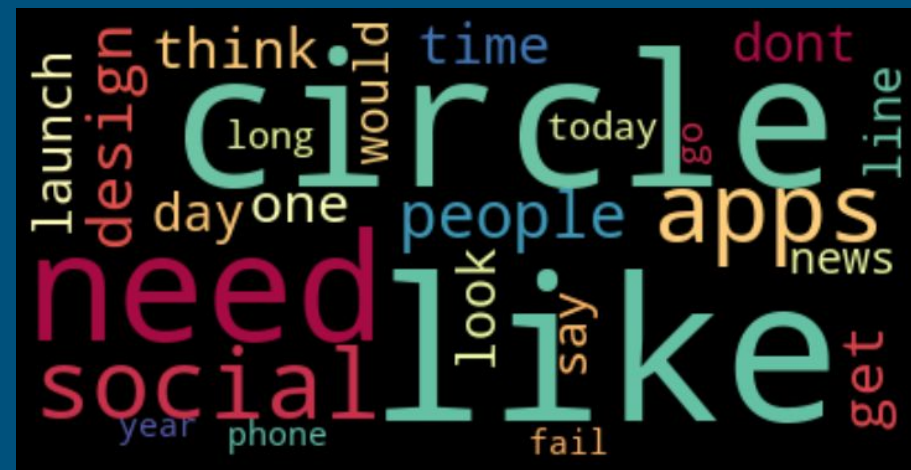
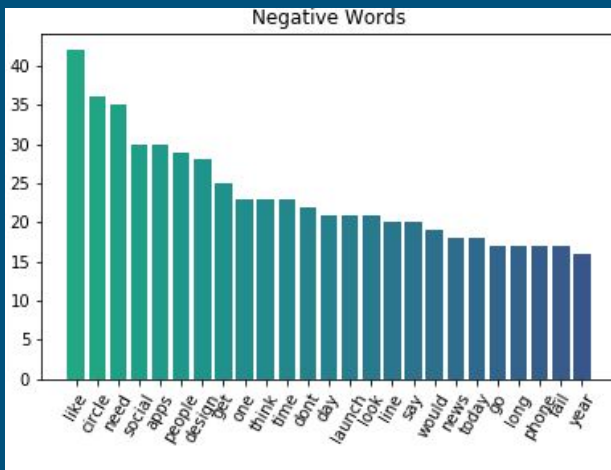
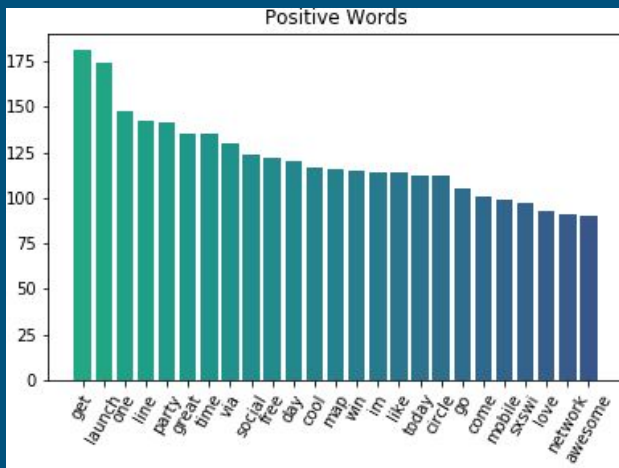
Data hasn't been updated since 2015.

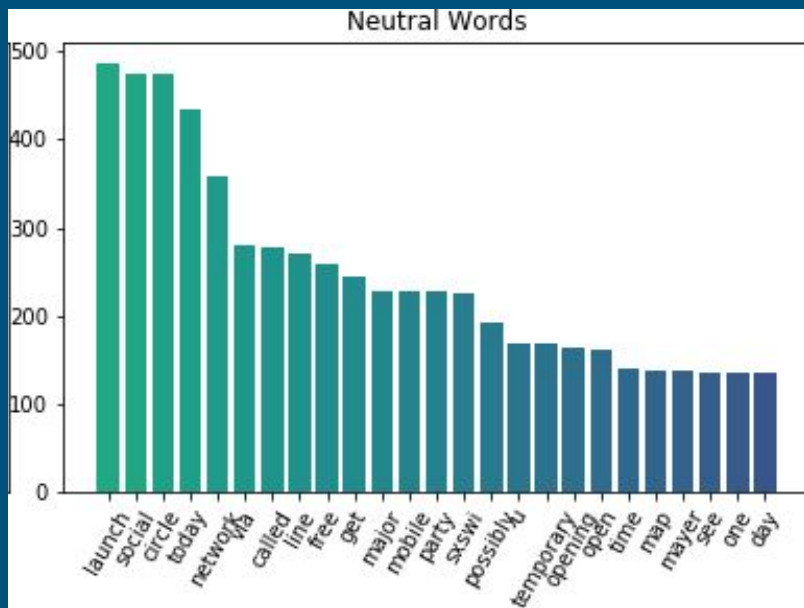
571 negative tweets, 156 I can't tell, 5389 no emotion towards brand or product, 2978 positive emotion

	tweet	product	emotion
count	9092	3291	9092
unique	8995	9	4
top	rt mention google to launch major new social n...	iPad	No emotion toward brand or product
freq	11	946	5388

# EDA

—





# Count Vector and TF-IDF with Random Forest

Evaluations for test:

```
[[ 25  78 10]  
[  8 974 102]  
[  3 349 239]]
```

		precision	recall	f1-score	support
No emotion toward	Negative emotion	0.69	0.22	0.34	113
	brand or product	0.70	0.90	0.78	1084
	Positive emotion	0.68	0.40	0.51	591
accuracy				0.69	1788
macro avg		0.69	0.51	0.54	1788
weighted avg		0.69	0.69	0.66	1788

Evaluations for train:

```
[[ 456   1   0]  
[  2 4299   3]  
[  0  33 2354]]
```

		precision	recall	f1-score	support
No emotion toward	Negative emotion	1.00	1.00	1.00	457
	brand or product	0.99	1.00	1.00	4304
	Positive emotion	1.00	0.99	0.99	2387
accuracy				0.99	7148
macro avg		1.00	0.99	0.99	7148
weighted avg		0.99	0.99	0.99	7148

Count Vector

Evaluations for test:

```
[[ 19  85   9]  
[  6 965 113]  
[  4 341 246]]
```

		precision	recall	f1-score	support
No emotion toward	Negative emotion	0.66	0.17	0.27	113
	brand or product	0.69	0.89	0.78	1084
	Positive emotion	0.67	0.42	0.51	591
accuracy				0.69	1788
macro avg		0.67	0.49	0.52	1788
weighted avg		0.68	0.69	0.66	1788

Evaluations for train:

```
[[ 456   1   0]  
[  2 4298   4]  
[  0  32 2355]]
```

		precision	recall	f1-score	support
No emotion toward	Negative emotion	1.00	1.00	1.00	457
	brand or product	0.99	1.00	1.00	4304
	Positive emotion	1.00	0.99	0.99	2387
accuracy				0.99	7148
macro avg		1.00	0.99	0.99	7148
weighted avg		0.99	0.99	0.99	7148

TF-IDF

# Conclusion

---

Conclusion shows that we can predict the sentiment of apple and google products with a 69% amount of accuracy.

This shows that we can accurately predict other approval rates for new products in the tech world.

# Next Steps

---

Use word embedding to more accurately predict emotion using context from tweets and not just word usage

Xgboost the models to refine the models to improve accuracy.

EDA to improve the accuracy of model by reducing tweets if all words contained are unique.





# Thank You

<https://github.com/asherkhan7/Tweet-Analysis-NLPgit>