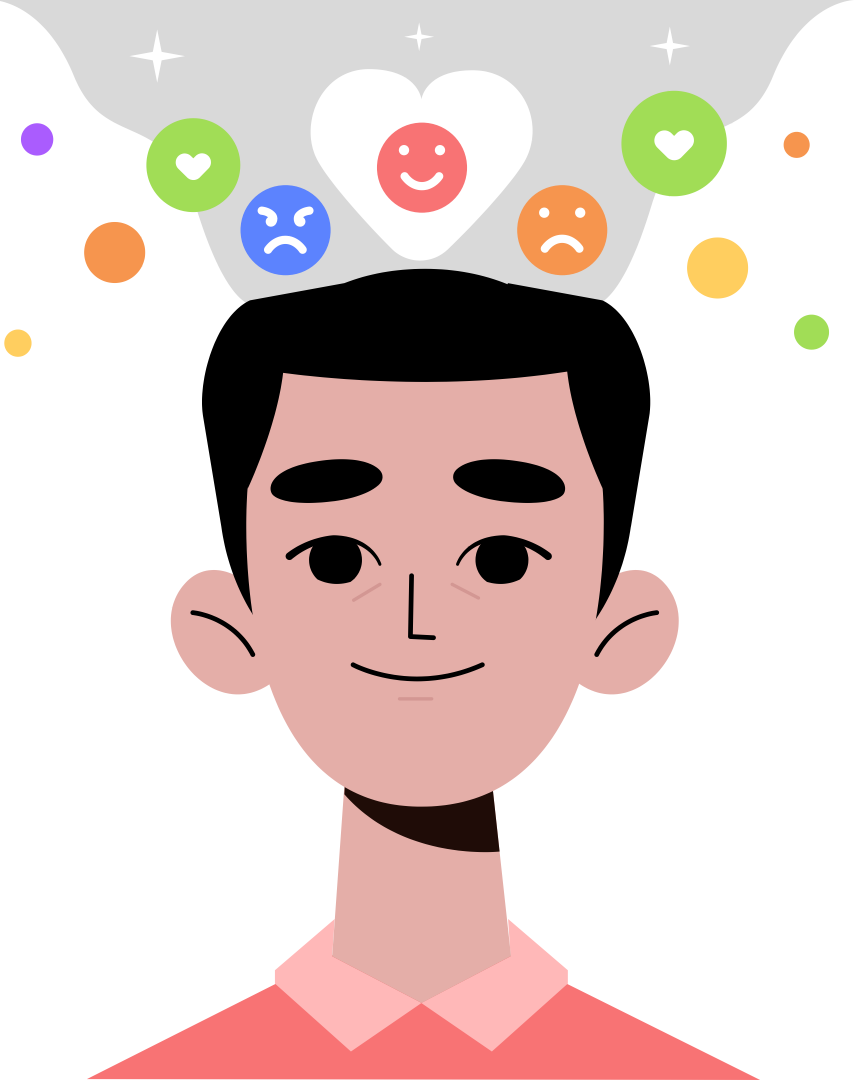


Twitter Sentiment Classifier

Gaining actionable insight from social media data

Outline

Business Problem

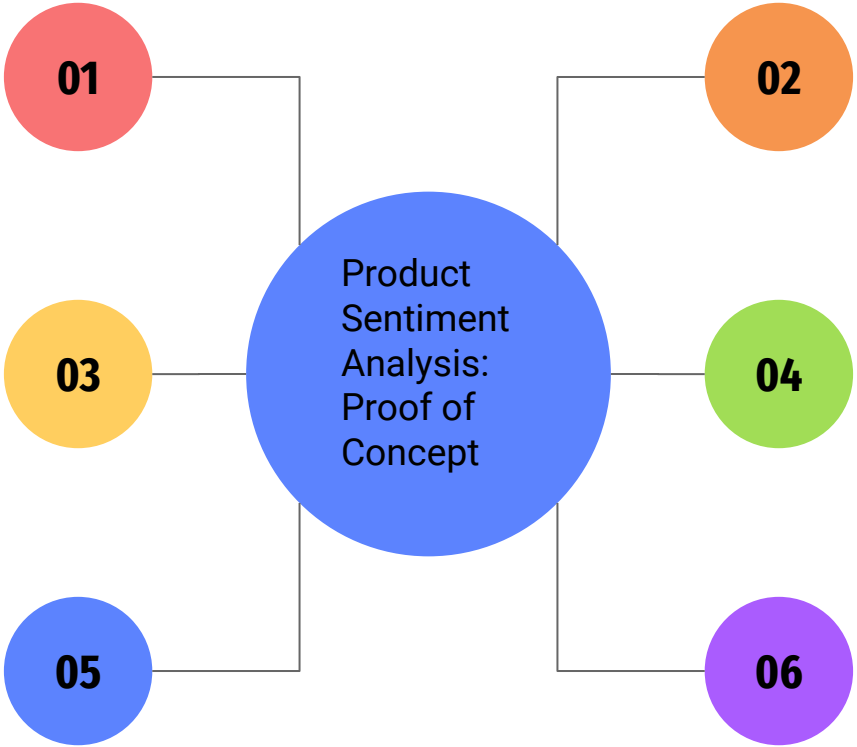
Process and classify twitter text data to gain insights on a brand and associated products.

Methods

Methods for gaining insight from twitter data and for building a classifier

Conclusions

Conclusions?



Data

Labeled twitter data relating sentiment to brands and products

Results

Evaluation of model performance and insights gained from test models

Future Work

Future Work?

Business Problem

01

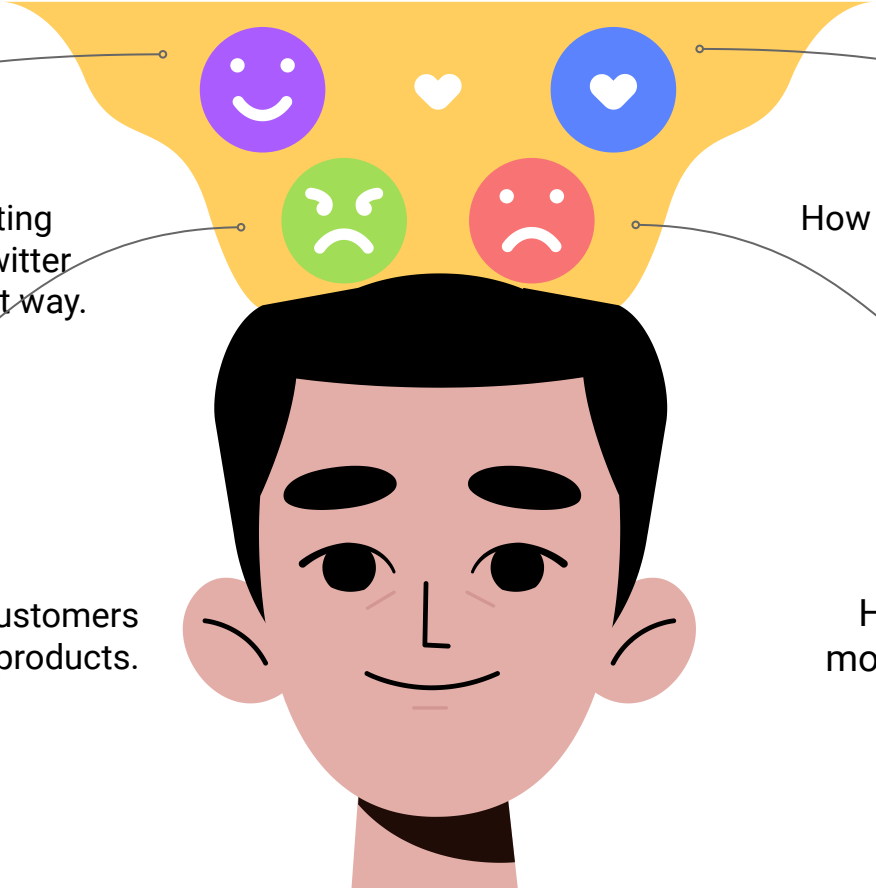
Twitter Data

explore options for generating actionable insights from twitter text data in a more efficient way.

02

Proof of Concept

How could machine learning be deployed to help?



03

Sentiment

interested in what their customers feel about their products.

04

Balanced Accuracy

How well can we trust the model to make predictions?

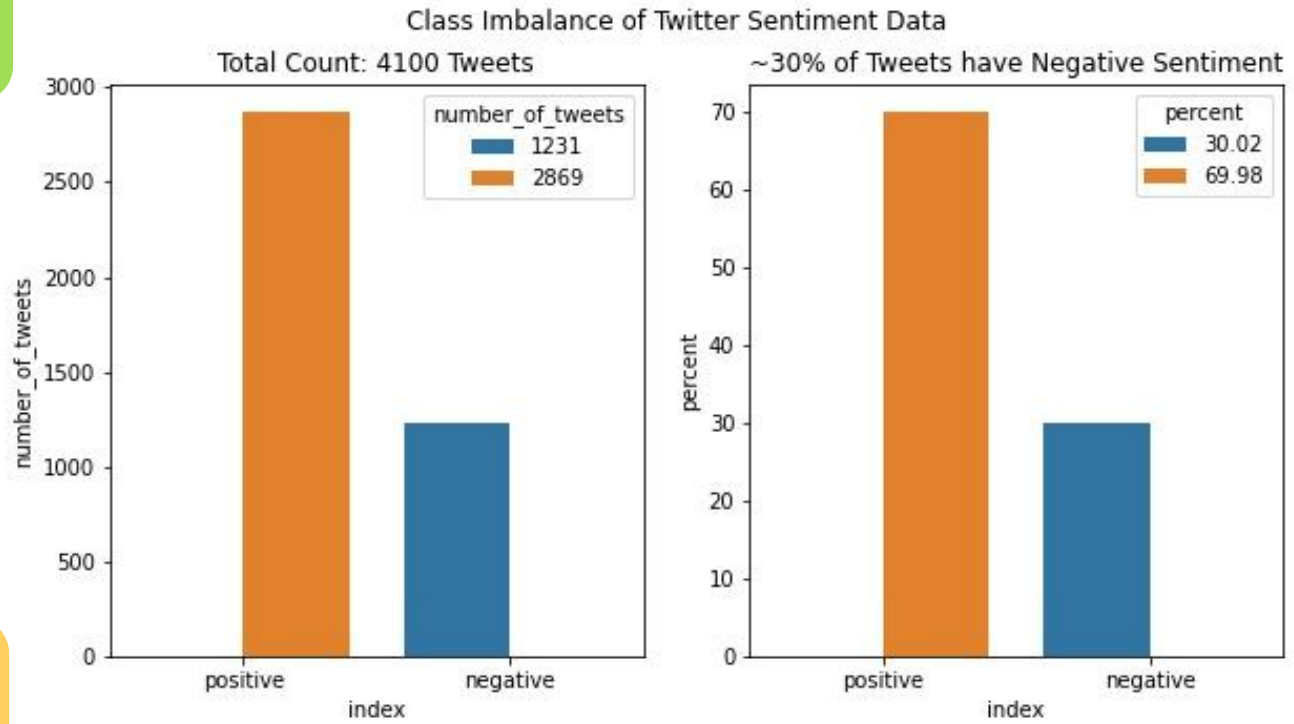
The Data

Crowdsourced labeled tweets from 2011 SXSW event (Apple/Google)

Labeled Twitter Data

Sourced from CrowdFlower

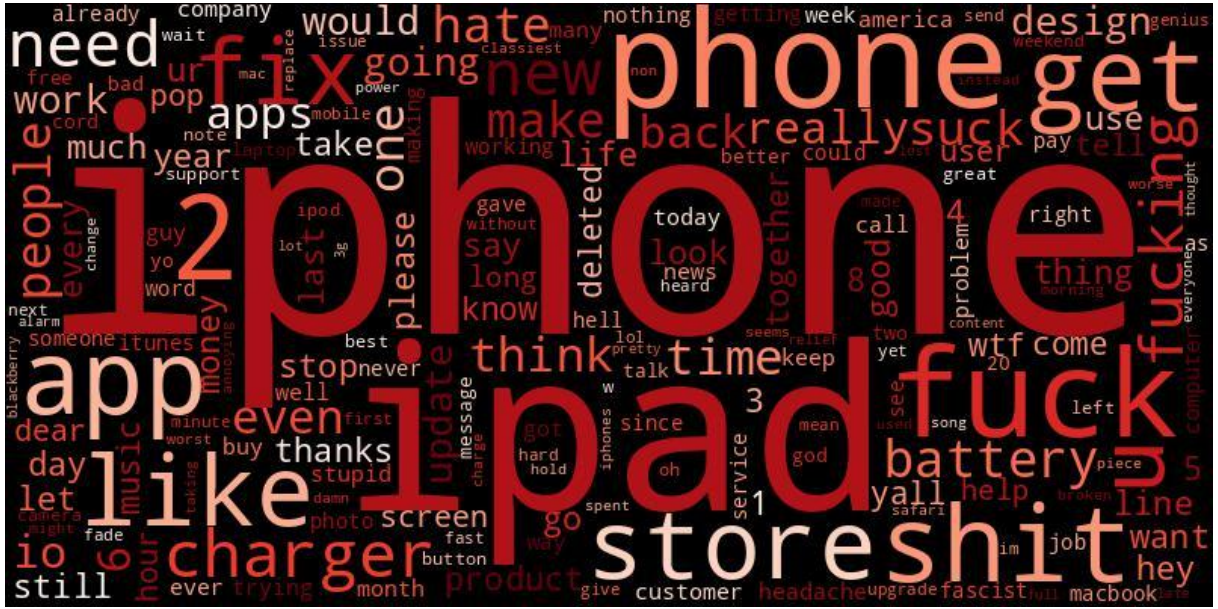
Additional sentiment labeled data available from CrowdFlower (Apple)



Methods: Natural Language Processing

Word Clouds

Use word frequencies to construct quick visualization of what customers do not like apple a company or product



PMI Scores

Calculate Pointwise Mutual Information to create bigrams, trigrams, and quadrgrams

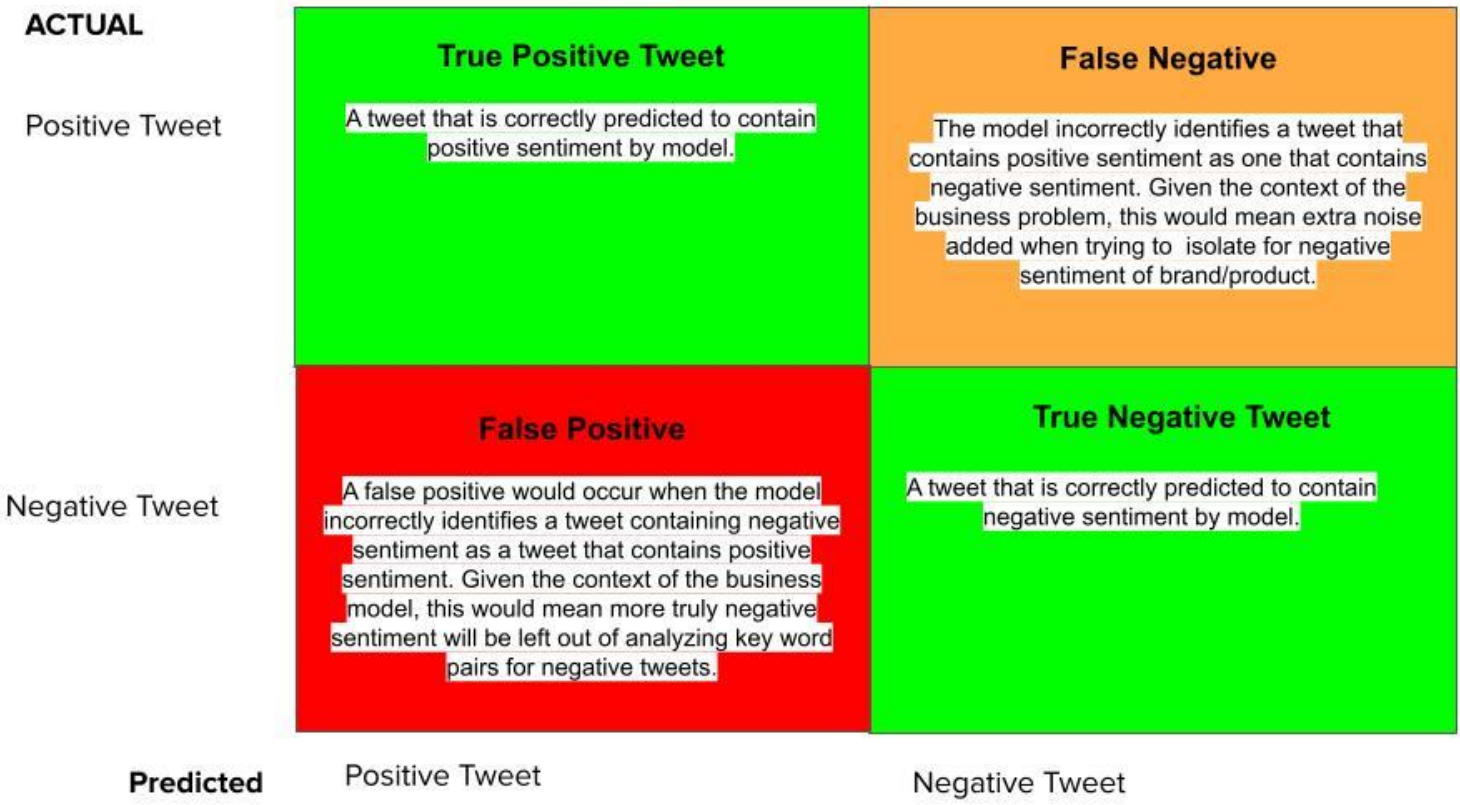
	word_triplets	PMI
0	(heat, million, sun)	21.286869
1	(button, heat, million)	20.412400
2	(classiest, fascist, company)	18.129017
3	(back, button, heat)	17.876347
4	(fascist, company, america)	17.851483
5	(news, apps, fade)	16.106960
6	(ipad, design, headache)	13.715867
7	(ipad, news, apps)	12.436422
8	(ipad, back, button)	12.298918
9	(iphone, battery, life)	10.537000

Methods: Classification Evaluation

Key Metric: Balanced Accuracy

average of recall obtained on each class

Model Performance Matrix



Results: Balanced Random Forest Classifier

81%

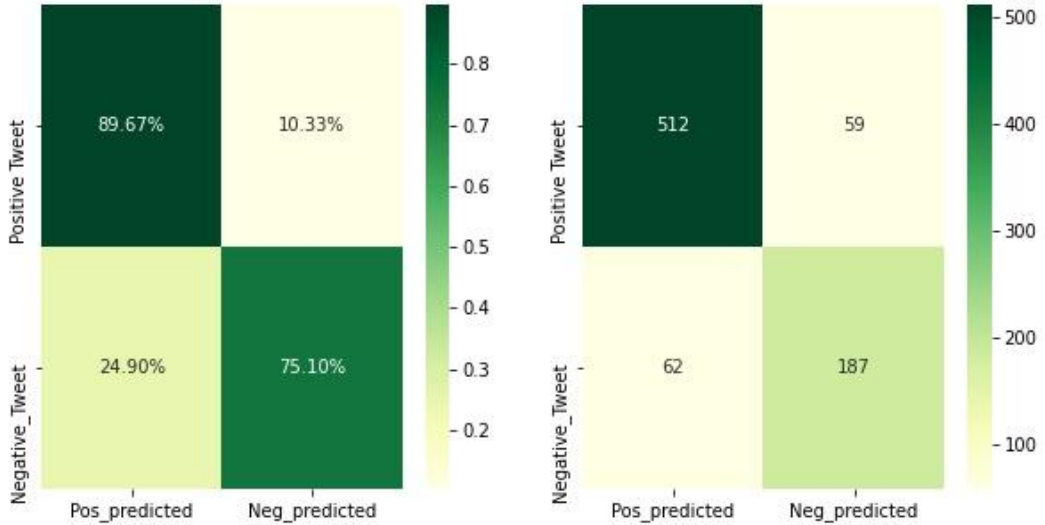
Balanced Accuracy

Despite being red, Mars is a cold place

Tuned

Hyperparameters

Tuned for balanced accuracy with gridsearchCV



See the appendix for more model results. The one shown here is the best performing supervised model.

Results: BERT_base Sentiment Classifier

~87.3%

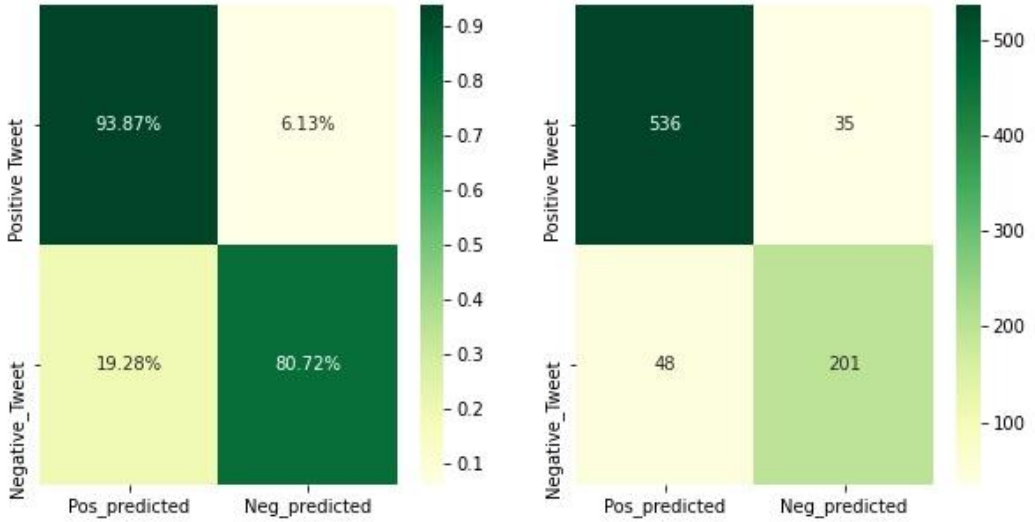
Balanced Accuracy

Despite being red, Mars is a cold place

pre-trained

Convoluted network

Transfer learning takes advantage of gigantic model



See the appendix for more model results. The one shown here is the best performing BERT model.

Conclusions and Future Work

Either classifier could be used to predict sentiment on new brand-centric social media data for the company's own products or that of a competitor

- Use the BERT classifier to predict the sentiment on new unlabeled twitter data filtered for product or brand of interest (Apple/Google) from another source to find more actionable insights to further proof of concept.

- Use the BERT classifier to predict the sentiment on new twitter data to help balance existing dataset and retrain the other models.

- leverage a state-of-the-art early stopping algorithm (ASHA) using Ray Tune and PyTorch

Thank You

Email
ddey2985@gmail.com

Github
Project Link

Blog
Blog Link

Thanks
Feedback Welcome