Using Machine Learning to Detect Product Sentiment in Tweets

A proof-of-concept

By Jessica Miles
June 20, 2021

Agenda

- Challenge statement: What and why
- My approach
- Tweet dataset overview
- Machine learning results
- Conclusions from POC

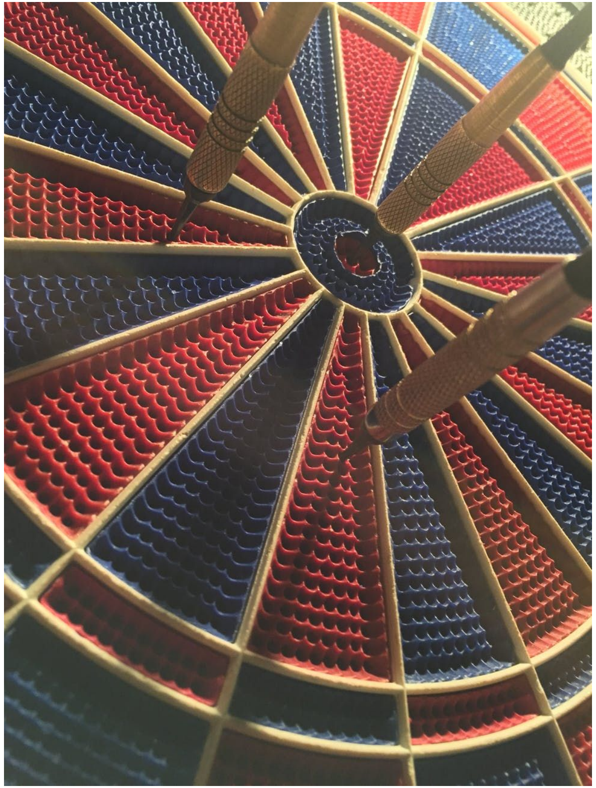
Photo by cottonbro from Pexels

The challenge: Understanding consumer sentiment

- Voluntary product reviews are often highly polarized [1]
- Analysis of support tickets or customer complaints may highlight areas for improvement, but miss what works well
- It can be difficult to design surveys that avoid response bias
- Publicly available data such as social media have far too much data to be manually reviewed

Can machine learning help?

- Can we use machine learning to separate tweets which contain positive or negative sentiment towards a brand or product from tweets which do not?
- What actionable insights could a machine learning model provide to a company interested in who wants to understand the factors driving both positive and negative sentiment?



My approach - A proof-of-concept

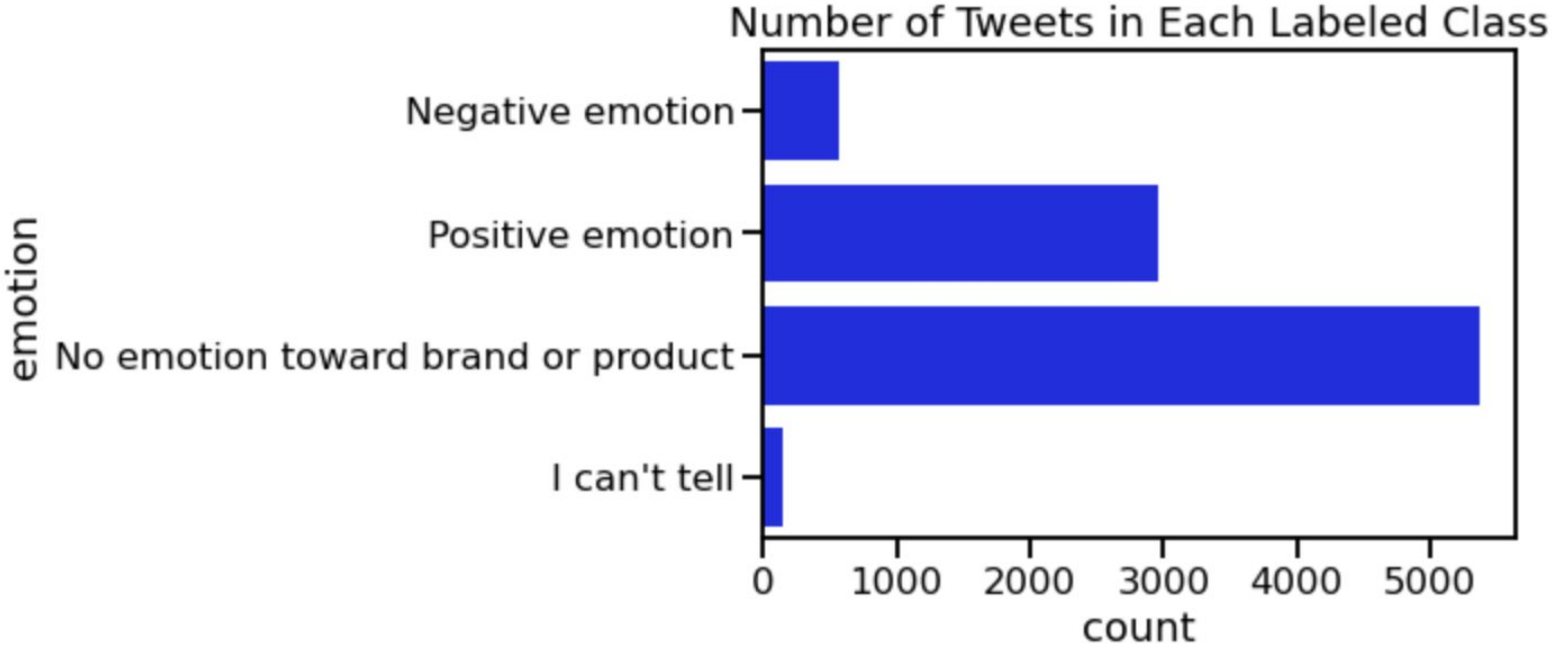
- Trained simple machine learning models on a set of tweets posted during SXSW
  - Models included Naive Bayes, Random Forest, and Logistic Regression
  - Simple models for POC, and performance can serve as a baseline to evaluate more complex models
- Evaluated model classification performance compared to human classification
- Extracted predictors of sentiment from highest-performing models to show what insights could be gleaned

Data used for analysis

- ~9,000 tweets from SXSW, most of which contain mentions of Apple or Google products or brands [2]
- Tweets were originally classified by humans into the following categories:
    - **No sentiment** towards a brand or product (or neutral emotion towards a brand or product)
    - **Positive sentiment** towards a brand or product
    - **Negative sentiment** towards a brand or product

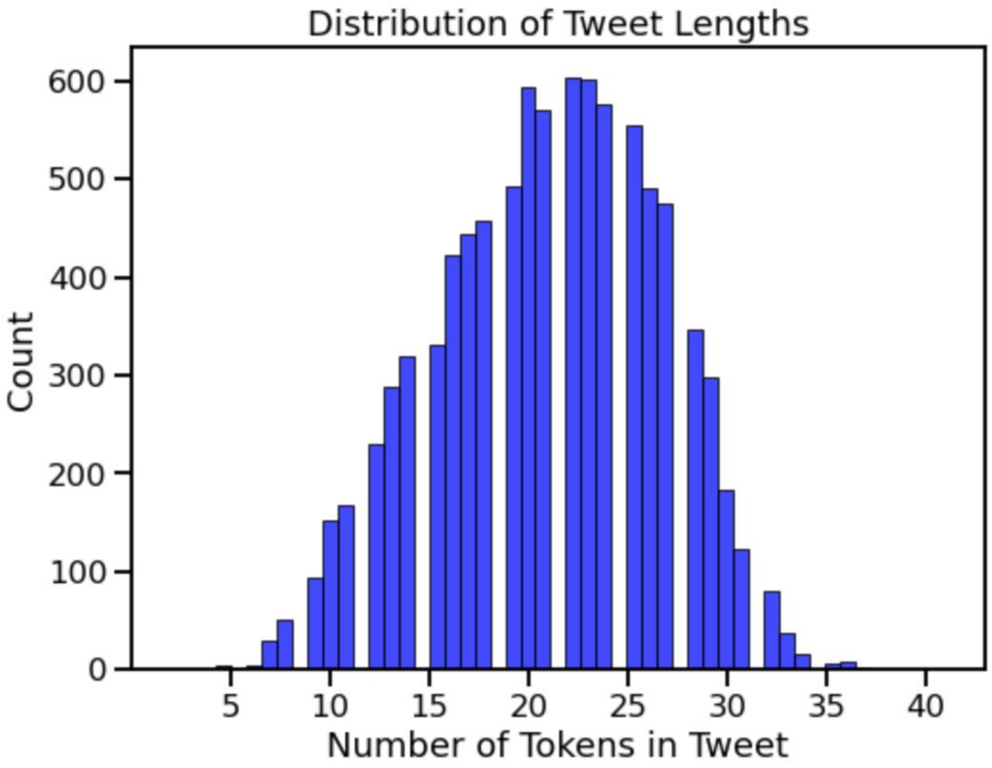
Tweet Dataset Overview

Class Distribution



Tweet Lengths (in word tokens)

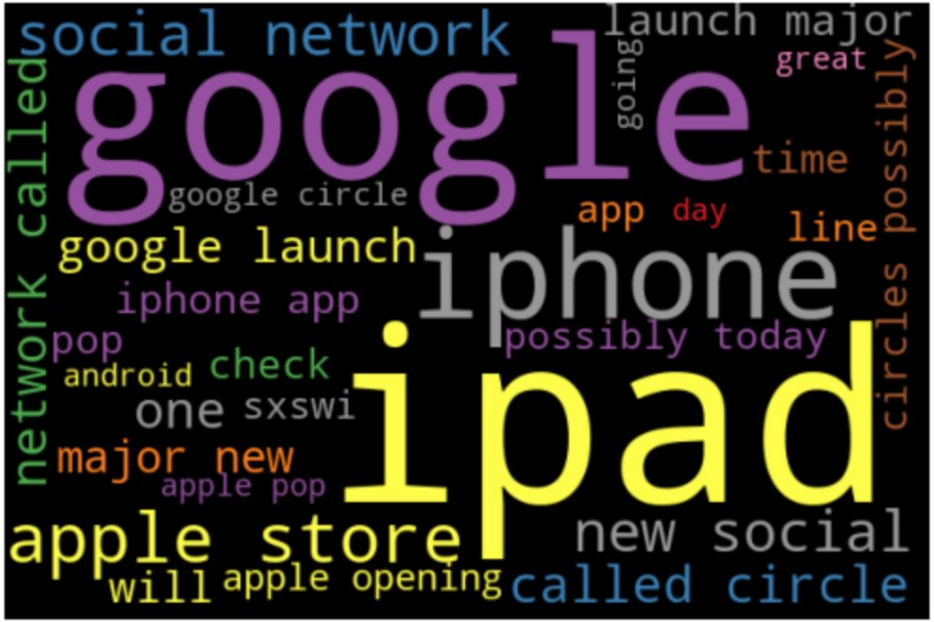
- Before removing stopwords, punctuation, and @mentions
- Most tweets are between 20 and 25 words



Word Frequencies - Entire Corpus

- Removed references to SXSW
- Product and brand-related words are very common

Word Cloud for Entire Corpus
("SXSW" and common stopwords removed)



Word Frequencies - Per Sentiment Class

No Sentiment Toward a Brand or Product

Positive Sentiment

Negative Sentiment

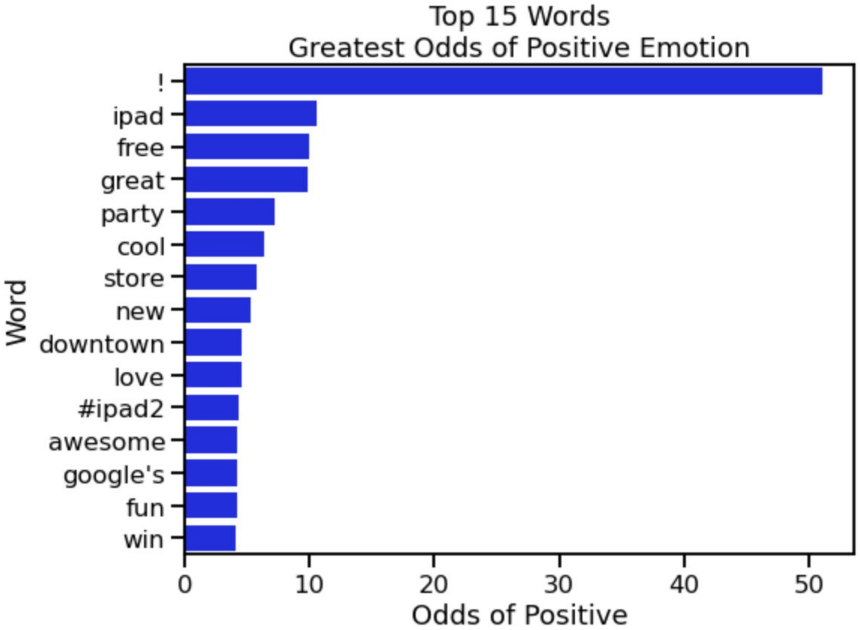
Machine Learning Results

Model Performance - Summary

- A simple multi-class model can achieve **~60-65% balanced accuracy** classifying Positive/Negative/No emotions on unseen test data
- A simple binary model can achieve **75% balanced accuracy** separating Positive from Negative emotions
- Both of these are **significantly better than random guessing**
- Even **humans only label only about 80% of data accurately** (in agreement with each other), so our models aren't much worse

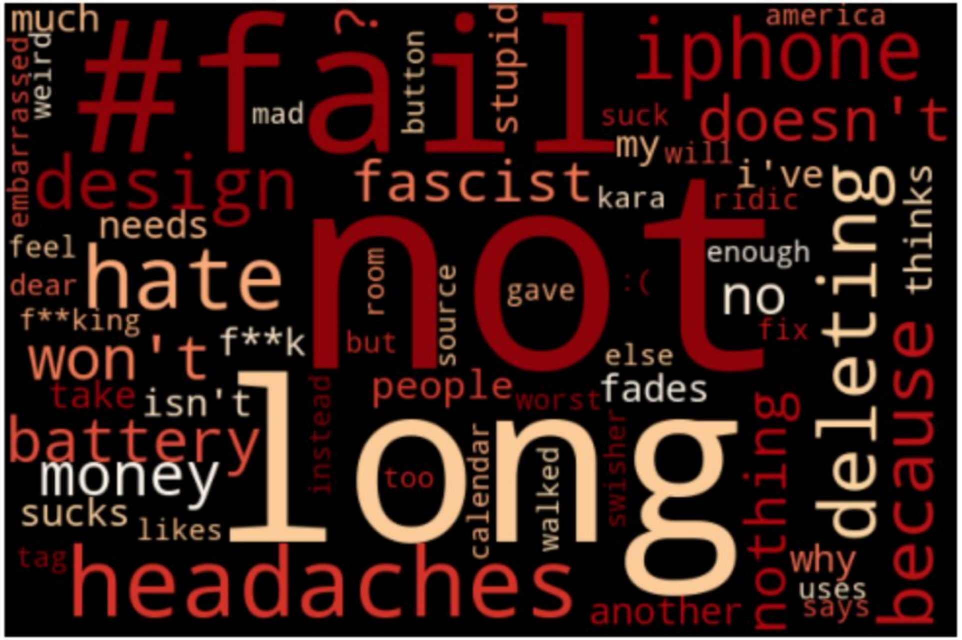
Insights about Positive Tweets

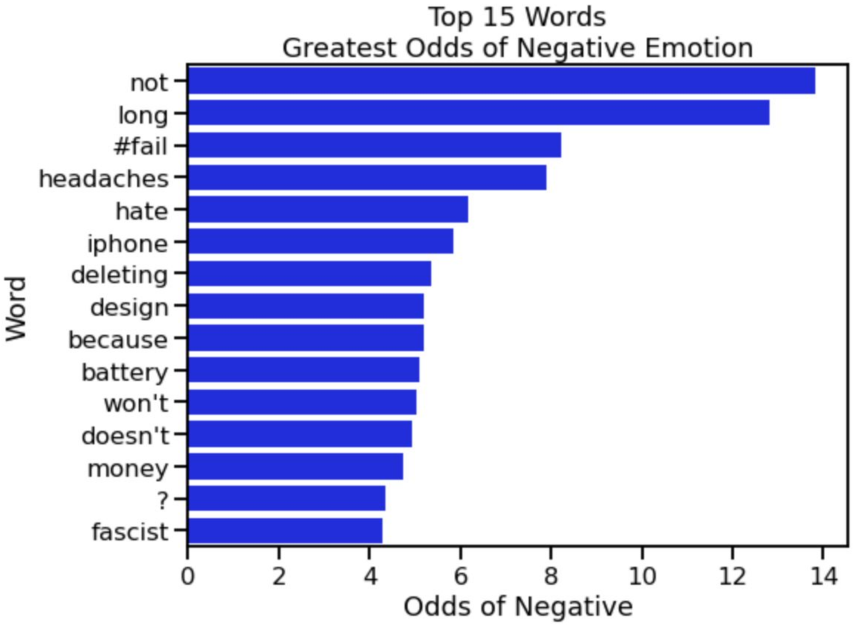
Top Predictors of Positive Emotion



Insights about Negative Tweets

Top Predictors of Negative Emotion





Conclusions from POC

Can simple models provide useful insights?

In general, yes!

- Simple models such as Logistic Regression CAN be trained to identify sentiment fairly accurately
- Simple models are also very interpretable, so we could easily provide samples of tweets with certain ngrams to SMEs for further understanding

Caveats:

- Supervised learning models can only be as accurate as the labeled data they're trained on, which will never be 100% accurate
- Training data should be selected carefully to maximize applicability on future unseen data. A specific goal, such as focusing on a particular product, may make this easier.

Proposed Sample Workflow using Simple ML Models

1. Company identifies goal of the sentiment analysis
2. Training tweets selected methodically with goal in mind, and humans label them (ideally SMEs)
3. Logistic Regression model trained on labeled data
    a. Multi-class model used to separate positive/negative from no sentiment
    b. Binary model used to provide key predictors of positive/negative sentiment
4. Key predictors of sentiment used to create samples of tweets for SMEs to review. They provide feedback about usefulness of insights.
5. Initial model tested on more tweets. As SMEs review new samples, they re-label as needed and corrected labels used to iteratively retrain model to improve performance

Recommended Next Steps

- Test recommended workflow where multi-class model feeds binary model
- Try out more complex models and preprocessing steps to see how much performance can be improved (simple models are performance baseline)
    - Convoluted Neural Network, SVM
    - Word embeddings
    - Pretrained vocabulary for sentiment

Thank you for reading!

For questions or comments, please contact:

Jessica Miles

jess.c.miles@gmail.com

Model Performance - Details

Multi-Class: Attempts to separate no sentiment towards brands and products from positive and negative sentiment

- Best model achieved **~60-65% balanced accuracy** across all classes on unseen test data
- **~30% better than guessing randomly** based on class distribution
- Performed about **equally well when predicting any class**
  - Confused Positive and No Sentiment more often than either of these with Negative

Binary: Attempts to separate positive from negative sentiment

- Best model achieved **~75% balanced accuracy** across both classes on unseen test data
- **~25% better than guessing randomly** based on class distribution
- **Better at predicting positive sentiment than negative**
  - Only ~10-15% of Positives misclassified as Negative
  - ~30-40% of Negatives misclassified as Positive