Apple & Google Sentiment Analysis

Samuel Mohebban

Raven Welch

The Problem

- Ideally customers are always satisfied
    - However this is rarely the case

- Twitter is a cornerstone for communication
    - Many customers use Twitter share their issues with companies

- Manual screening is costly and time-consuming

Why should you care?

- Twitter is easily accessible

- Conflict resolution can bring customers back

- Improves company image

Our Solution

Create an algorithm that can detect tweets with *Negative* sentiment

✓ Increased productivity for resolving customer conflicts

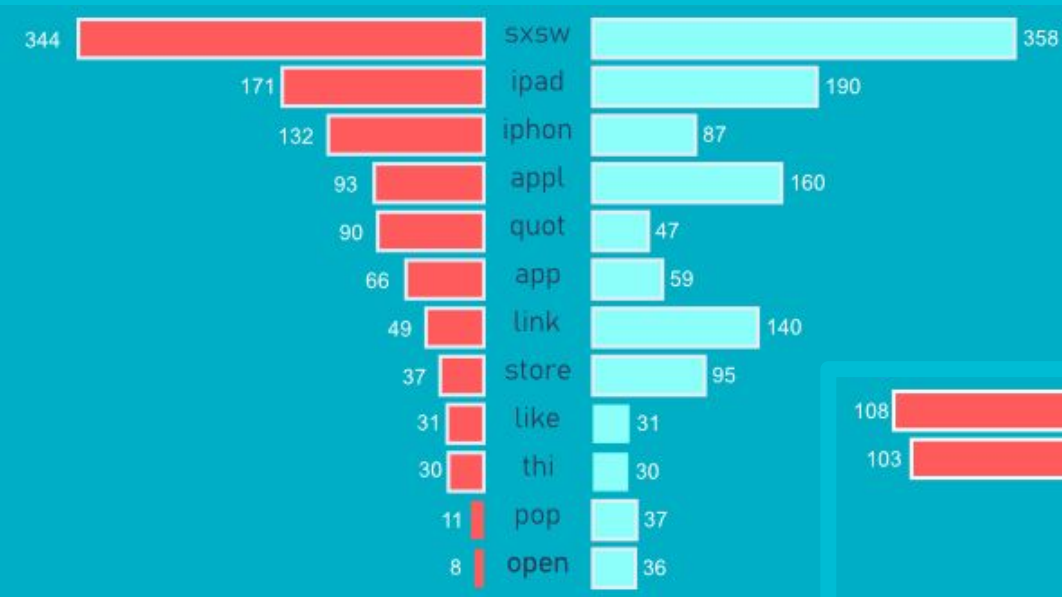✓ Automating sentiment screening saves money and improves customer

Data

- 9000 Tweets regarding Apple & Google Products collected by crowdflower on data.world

	No Emotion		Positive Emotion		Negative Emotion		Unknown

5389	2978	570

- Removed "No Emotion" and "Unknown" due to data imbalance

- Data was pruned to 600 Positive Tweets & 570 Negative Tweets

Top 12 Words in Negative and Positive Tweets

Our Process

- Cleaned the data to remove names, usernames, non alphabetical characters, and stop words such as "so", "a", and "we"

- Stemmed each word within our data
    - Ex: "Likes", "Liked", "Likely", "Liking" all become "Like"

- Applied KMeans clustering

- Split into Training and Test sets for modeling
    - Train → 994 (50% Positive, 50% Negative)
    - Test → 176 (50% Positive, 50% Negative)

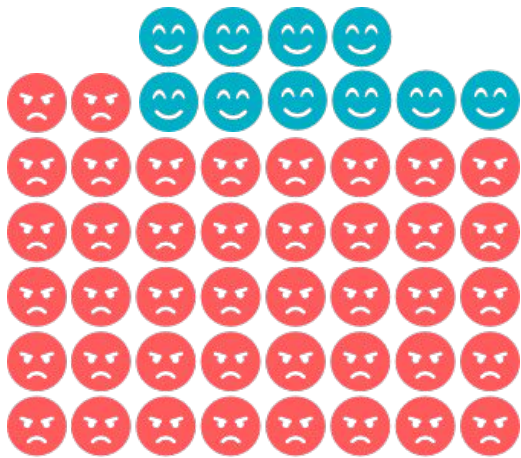- Fed the data into many different kinds of models and analyzed the results

KMeans & PCA Clustering
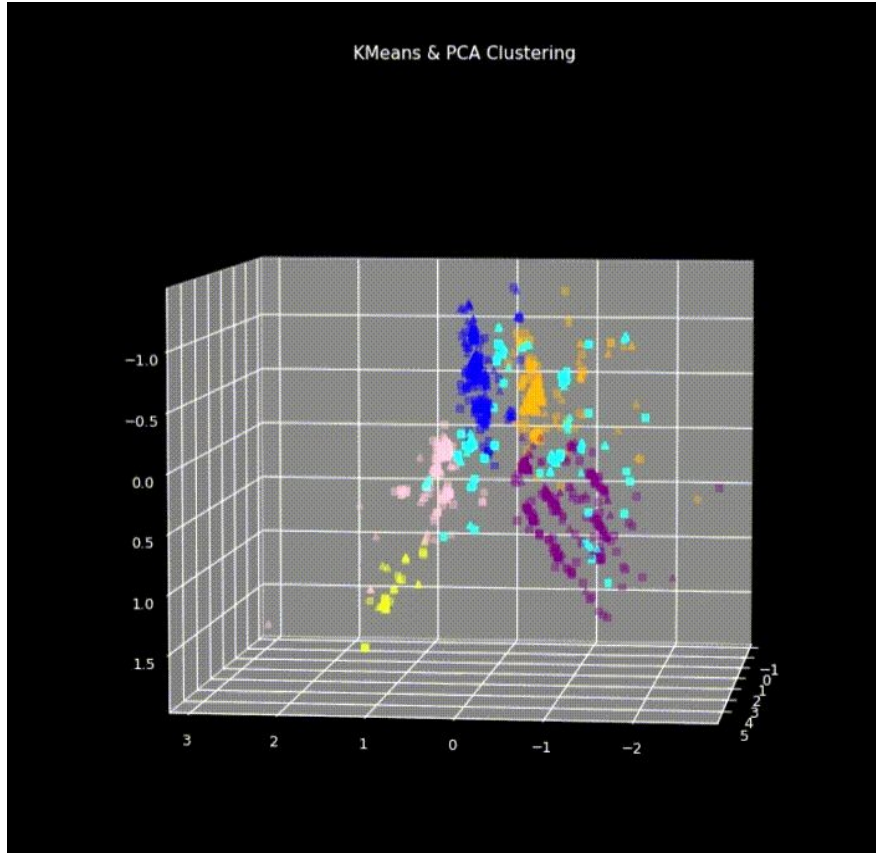
Model Performance

For every 100 tweets, our model…

- Correctly identifies 42 **Negative Tweets** as **Negative**

- Misidentifies 10 **Positive Tweets** as **Negative**

- Misidentifies 8 Negative Tweets as **Positive**

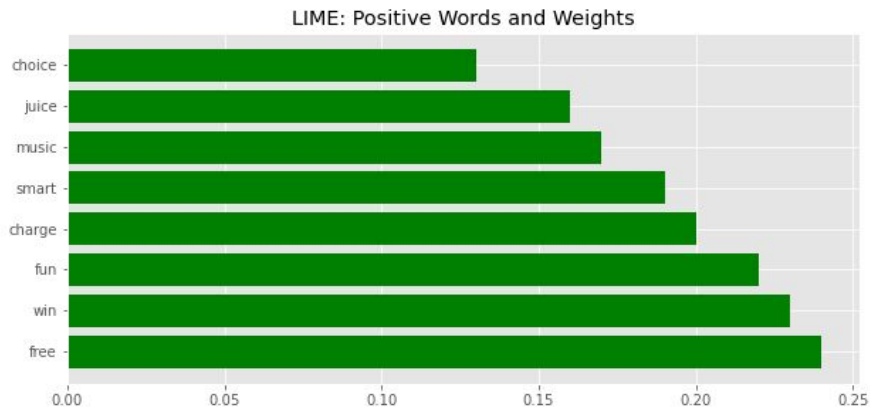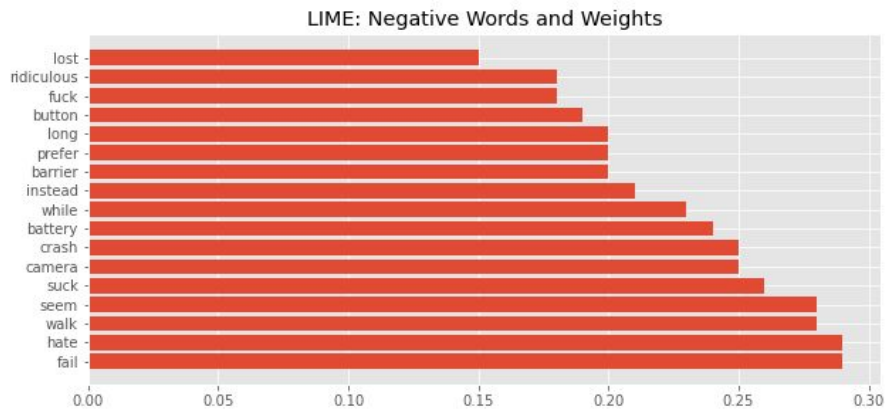- (Not Shown) Correctly identifies 40 **Positive Tweets** as **Positive**

Icons created by Font Awesome

Feature Importance

Counts for Words found in Each Model's Feature Importance

Local Interpretable Model-Agnostic Explanations (LIME)

LIME: Negative Words and Weights

LIME: Positive Words and Weights

What does this mean?

Recommendations

- Invest in more data collection

- Refine model to include internet slang

- Analyze what users are frustrated with

- Include the ability to detect neutral sentiment

Limitations

- Small & Unbalanced Dataset

- Data was from August 30, 2013 (~7 years ago)

Authors

- Samuel Mohebban
    - Samuel.MohebbanJob@gmail.com
    - [GitHub](#)
    - [LinkedIn](#)
- Raven Welch
    - RavenNHWelch@gmail.com
    - [GitHub](#)
    - [LinkedIn](#)