

“Tells” in Tweets: Developing a reliable Twitter sentiment analysis classifier for tech companies

Michael Burak

Problem statement:

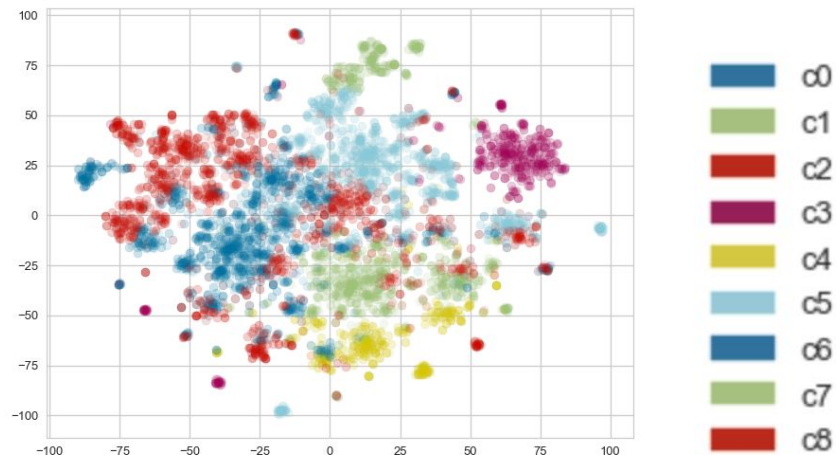
- Classifying tweets about your brand *reliably* is key.
- Twitter data is full of noise but can be very valuable.
- Developing a way to sift through the noise is key.

Sample Clusters:

0 : *ipad, sxsw, link, mention, apple, rt, the, design, new, line*

1 : *network, social, circles, called, launch, major, new, today, possibly, google*

2 : *apple, store, popup, sxsw, link, austin, ipad2, mention, line, open*



Business value:

- Reliable classifiers: good classifiers for the modern world.
- Through studying Apple and Google's wins and missteps, stand on the shoulders of giants.
- The data: Tweets about Apple and Google during SXSW 2011 and emotions expressed.

Methodology: an overview

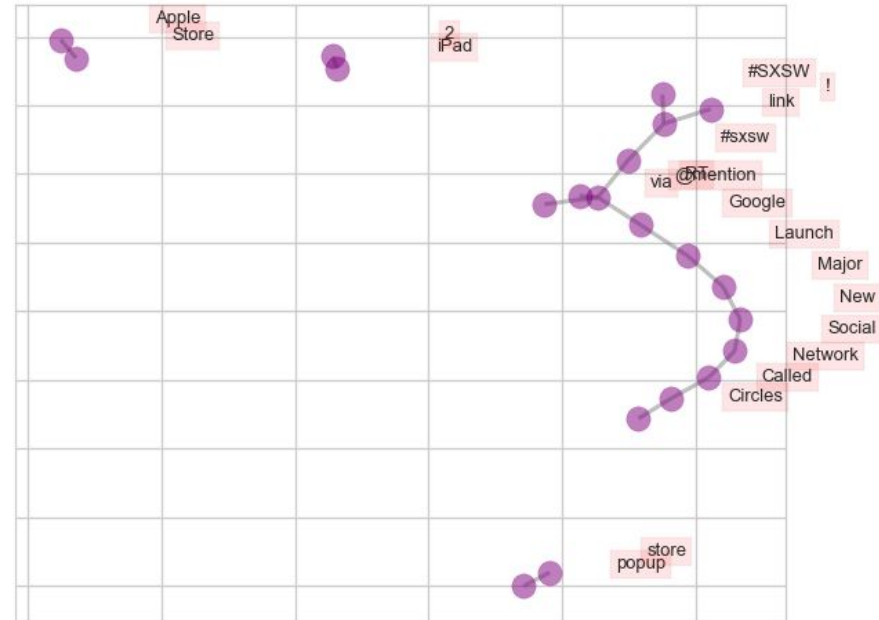
- Sentiment analysis: The study of emotions in text.
- Cohen's Kappa: Focus on reliability.
- Unsupervised learning: Finding patterns and comparison points.
- Naïve Bayes: Estimates probabilities, classifies well(with some tweaks.)

Business recommendation 1:

-Go big: capitalize on big events.

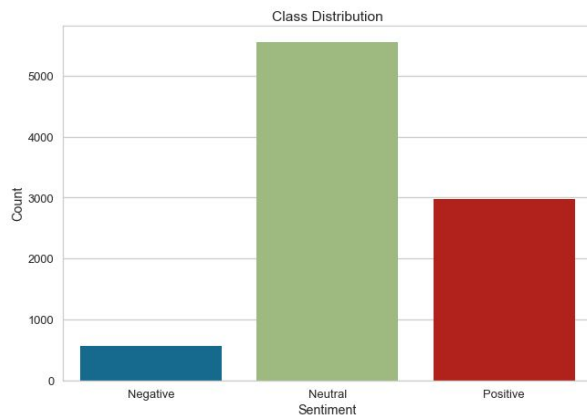
-Apple's popup store and iPad 2

-Google's social network "Circles"



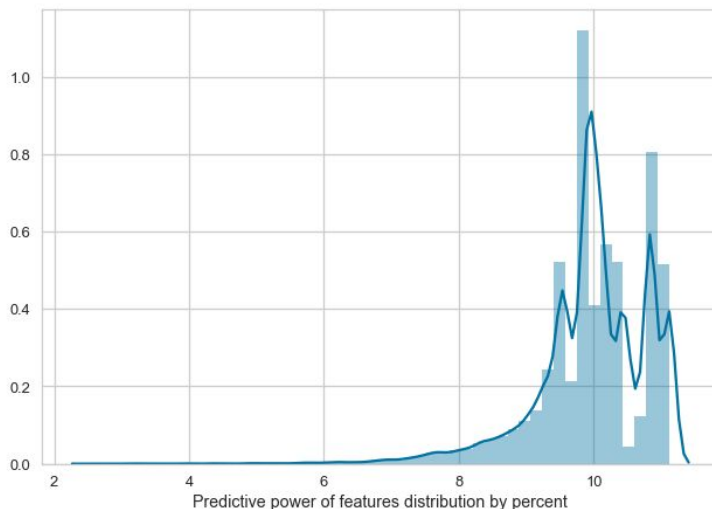
Business recommendation 2:

-Keep your data clean and representative: data integrity and imbalances' effects.



Business recommendation 3:

-Focus on the tangible and reliable: the iPad 2's predominance & results from models like ComplementNB with Cohen's Kappa for reliability. (Modern data is noisy.)



Future Work:

- Anomaly detection
- Deep Learning
- More Machines(Support Vector that is.)

Thank you!