Proof Positive

Can a machine find love on Twitter?

What did you want to know?

Public opinion about SXSW

What can we emphasize?

Do you want to know what went wrong?

Your data

~10,000 tweets were analyzed

The messages were graciously labeled and stored by an outside party: https://data.world/crowdflower/brands-and-product-emotions

Methods

Natural Language Processing

“Count Vectorization”, aka the easy method, worked best

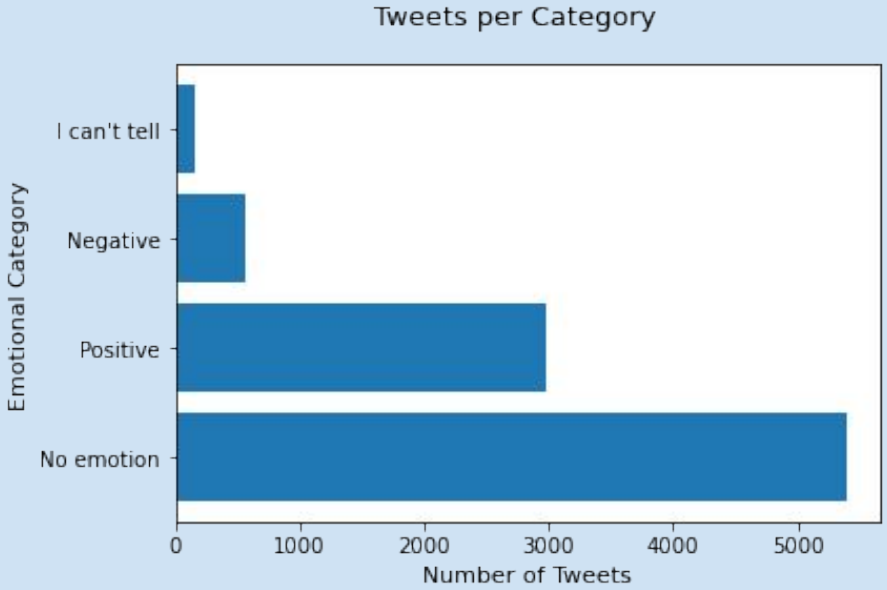
“Signs Point to Yes”

However, there were a couple key limitations

Your data = Proof of concept

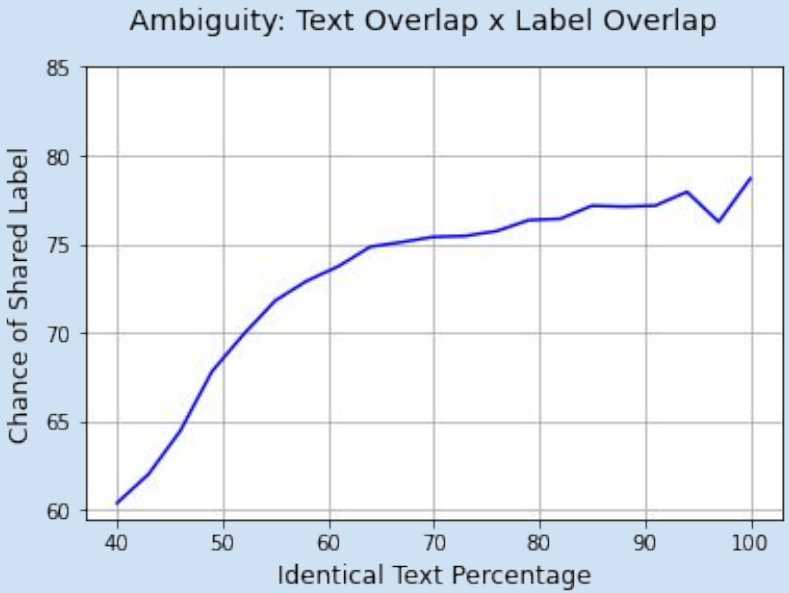
“My reply is no”

The good news: most people liked SXSW



“Reply Hazy, Try Again”

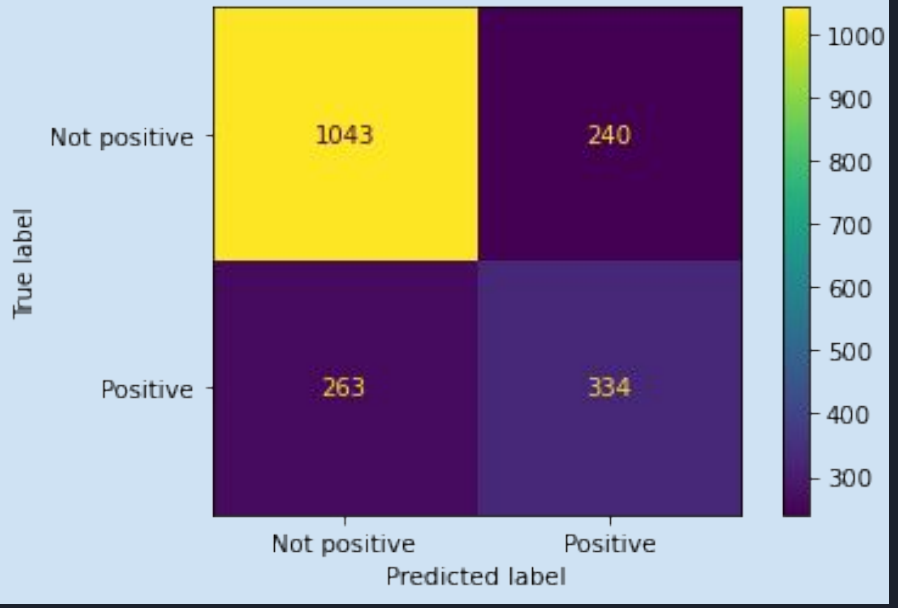
Even identical messages were not always categorized together



Results

The model achieved 73% accuracy on unknown data.

“Positive” = only 30% of the data, but the model correctly labeled nearly 60% of those tweets.



How can we improve? For your team:

Size of the data pool

Tweets as a medium

More reliability in categorization

How can we improve? For you:

See if the goodwill around Marissa Mayer still exists

Avoid new social media ventures, if at all possible

Potential Pitfalls

Resources spent labeling

Human error

Thank you

Feel free to reach me at:

Ferdinand.Beaman@gmail.com

or

www.linkedin.com/in/ferdinand-beaman

Credit to:
https://www.cleanpng.com/png-magic-8-ball-8-ball-pool-eight-ball-crystal-ball-8-710716/