Identifying disgruntled tweets

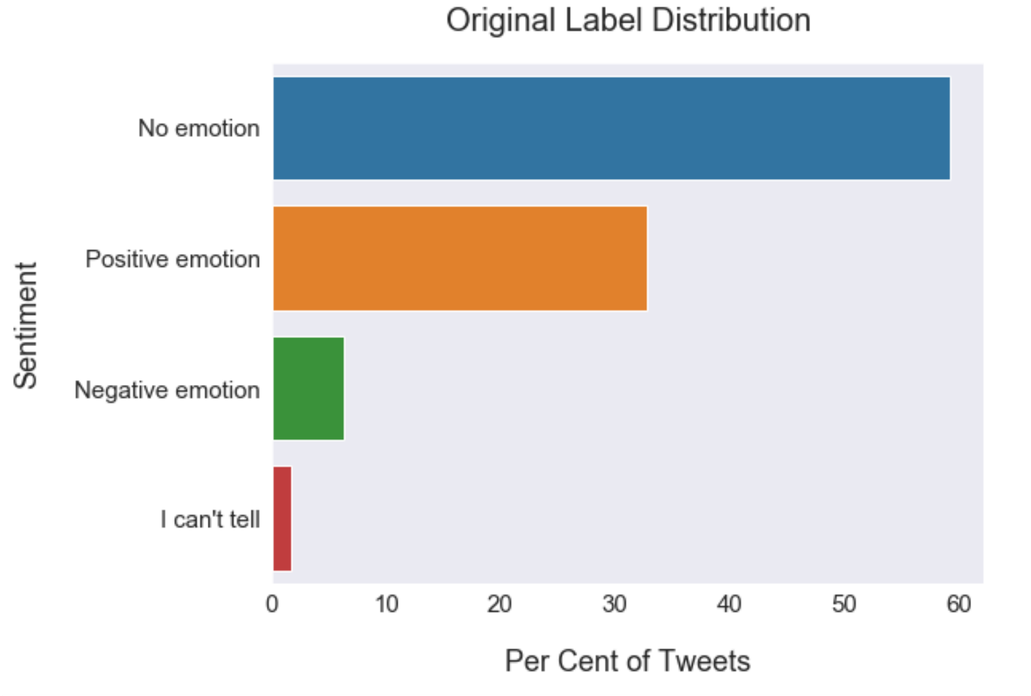
And how to increase Android's market share

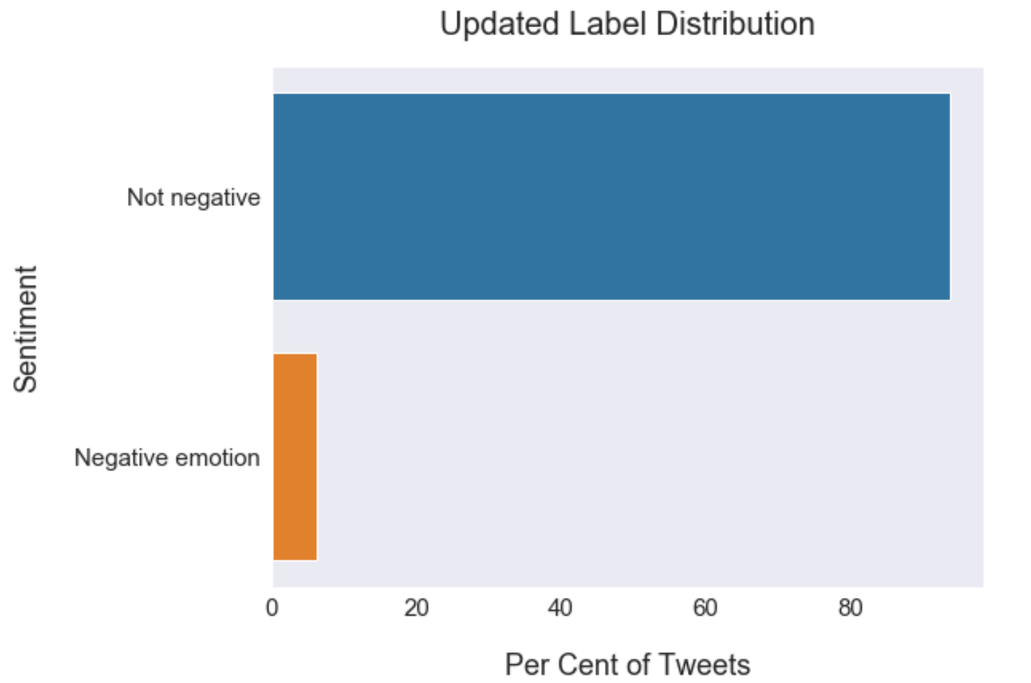
Business Problem

- Google wants to increase Android's share of the U.S. smart phone market
- Google is analyzing social media posts to find negative comments about mobile devices
- Analysts spend too much time reading through tweets that are positive or neutral
- Need a model to identify angry tweets that allows analysts to work at least twice as fast
- Need to retain as many angry / griping / complaining tweets as possible

Data

- Over 9,000 tweets about mobile devices and apps from 2011
- Each tweet has been rated as positive, negative or neutral
- Negative-sentiment tweets are only 6% of all tweets
- For this project, tweets are classified as negative-sentiment or not negative





Natural Language Processing Model

- Complement Naive Bayes model
- Count vectorizer
- Max 1,500 words
- 1-, 2-, and 3-word blocks
- Stop words list - 3 words only!
  - 'sxsw', 'mention', 'rt'

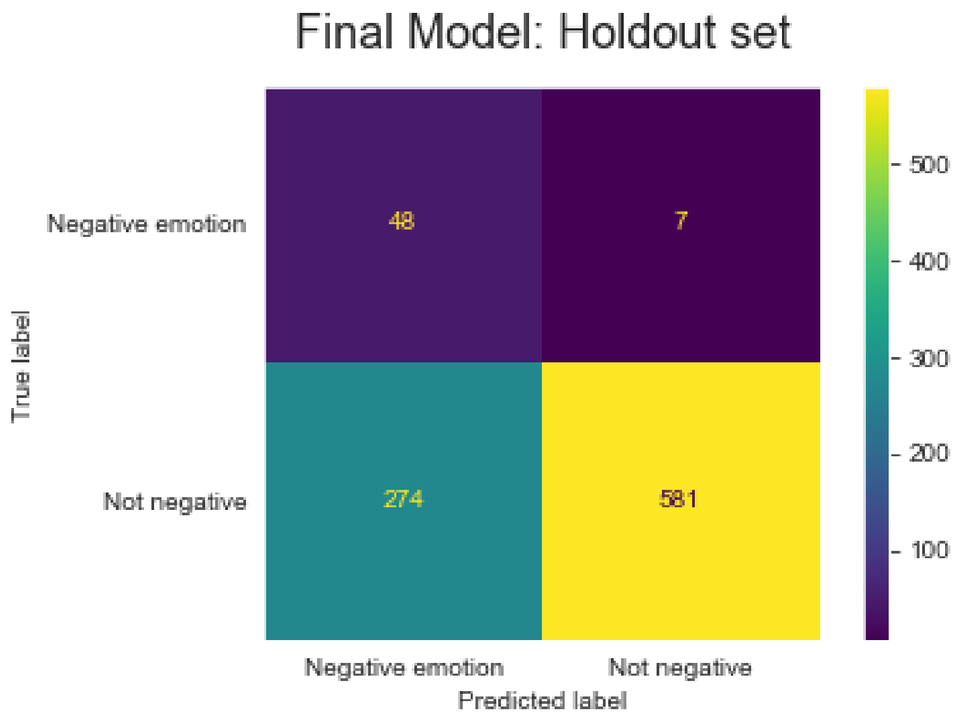
- 77 % of complaining tweets captured
- 14 % of tweets returned are complaints

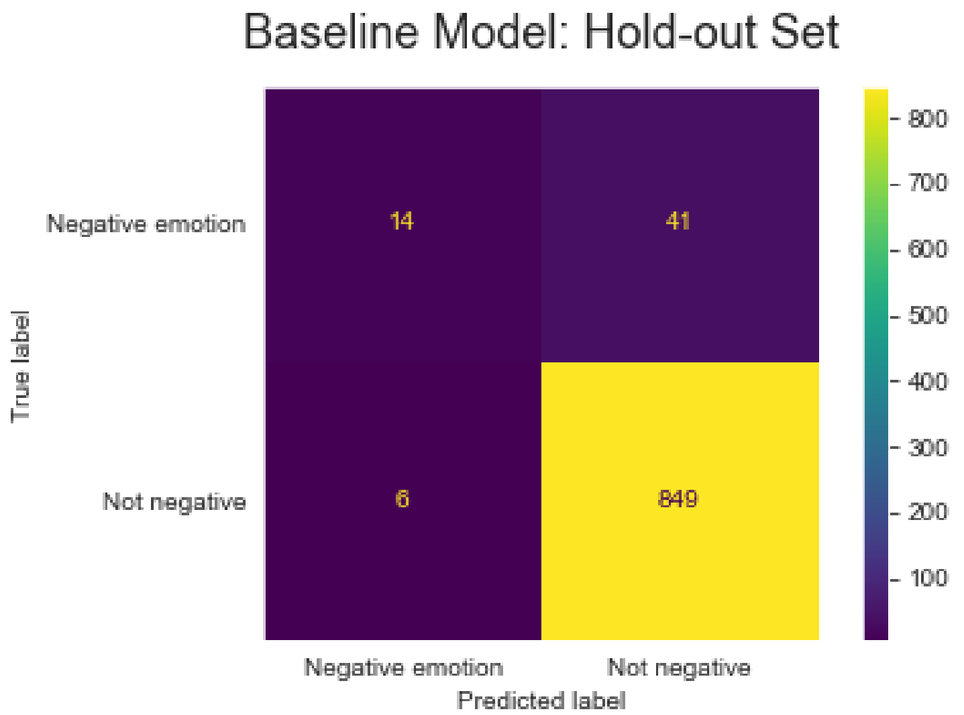
Results

Analysts can work over twice as fast, while sacrificing only 23% of the valuable information in griping tweets

Recall: 0.77 77% of complaining tweets were identified by the final model

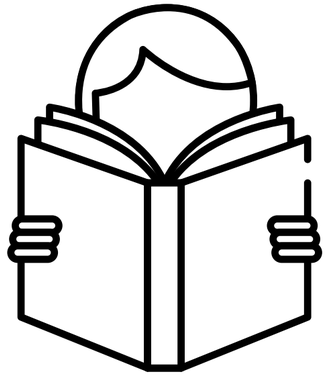
Precision: 0.14 14% of tweets returned by the final model were complaining





Conclusions and Future Work

Model effectively reduces the amount of time analysts spend reading tweets



For sentiment analysis of short posts, removing stop words may not make sense

“ ?! ”

Future Natural Language Processing Tools to test:

- Pre-trained word vectorizer
- Decision-tree based model
- Neural net

Thank you!

Questions?
Comments?

Lili Beit
Email: lilisbeit@gmail.com
github: @lilisbeit
LinkedIn: www.linkedin.com/in/lili-beit-33bb823/