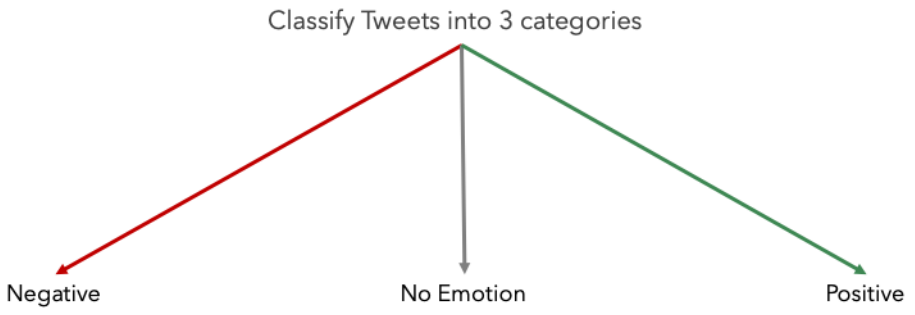

FLATIRON
SCHOOL
CLASSIFYING PRODUCT REVIEWS BY SENTIMENT
STEPHEN LANIER

PROBLEM STATEMENT



Technical challenge: evaluating multi-class classifier performance is a little trickier than binary, yes-no problems.

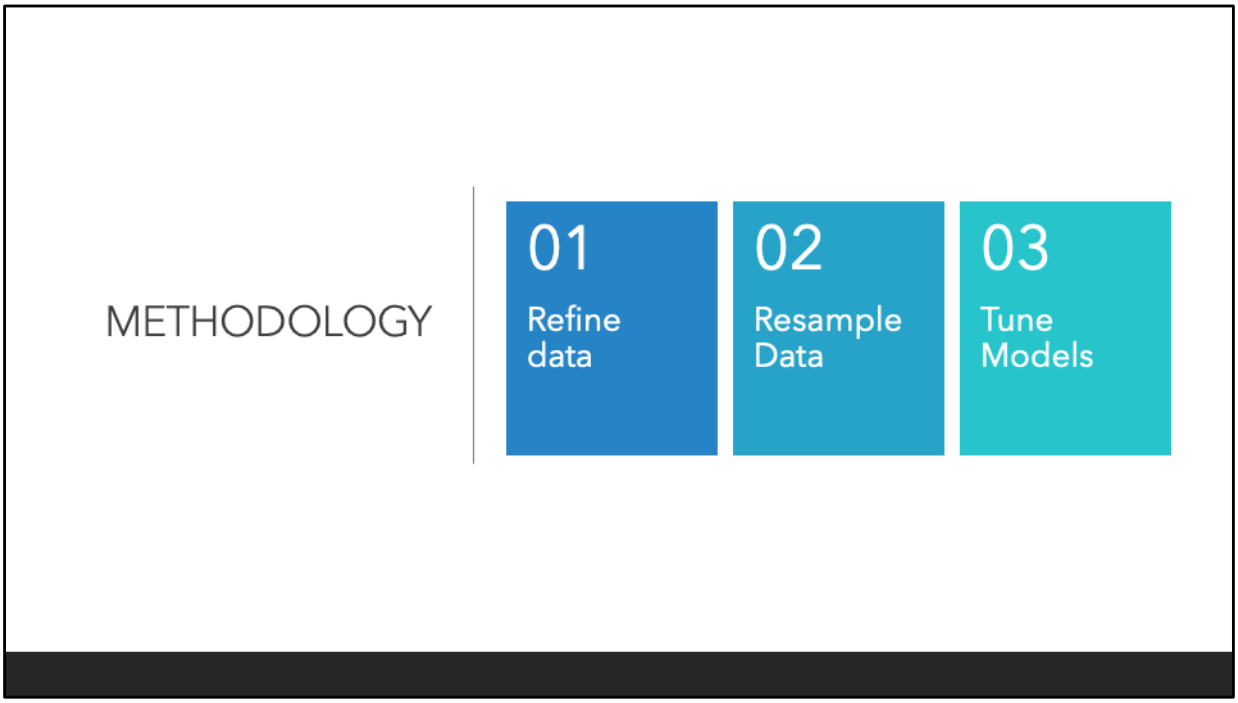
BUSINESS VALUE

Faster, deeper understanding of brand reception on web

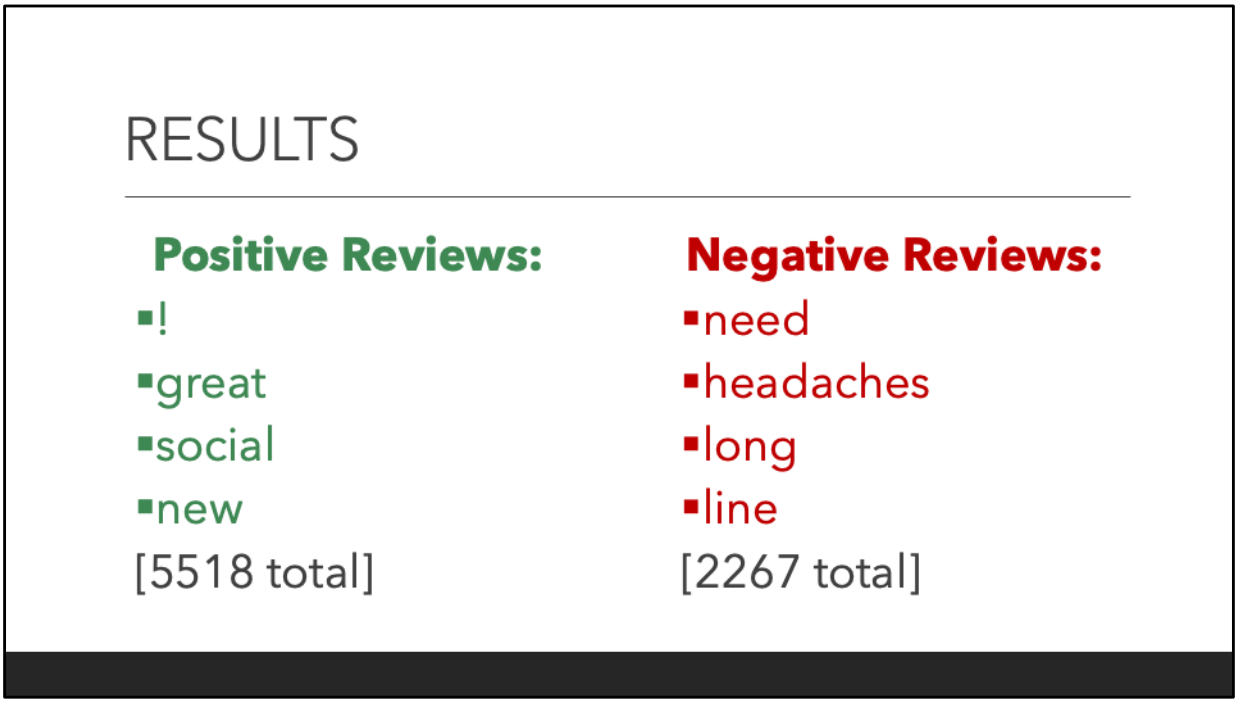
Brand	Product	% Positive	% No Emotion	% Negative
Google	Chromebook	40	50	10
	…	…	…	…
Apple	iPad	35	60	5
	MacBook Pro	45	55	0
	…	…	…	…
…				

Example implementation. Tweet data is broken down by brand and product.

Productionized, could feed in a large number of tweets and get a breakdown of percent positive, percent negative reviews by brand and product; by day, month, or year; or even by region, given more data.



1: split text data into a bunch of single word/character chunks; lemmatize; vectorize.
2: data is imbalanced with far more no-emotion reviews than positive or negative reviews.
3: choosing the right hyperparameters and engineering better neural networks.

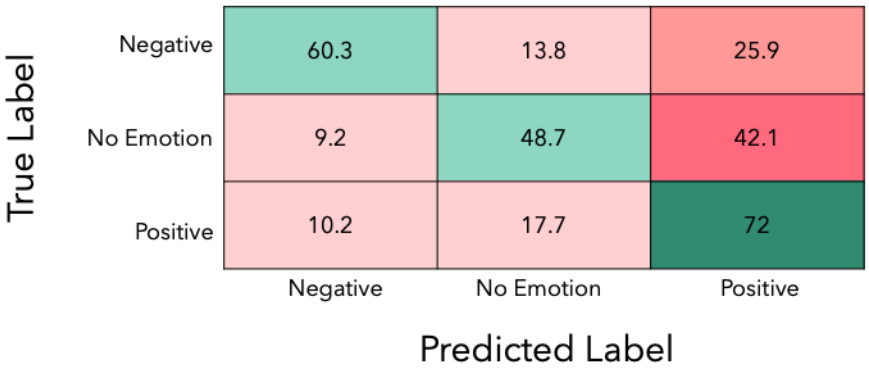


5518 unique lemmatized tokens from positive reviews, 2267 from negative reviews.

Model	Initial F1 Score	Best F1 Score
Neural Network	0.25	0.53
Logistic Regression	0.53	0.53
Random Forest	0.49	0.51
Support Vector Machine	0.47	0.48
Naïve Bayes Classifier	0.37	0.38

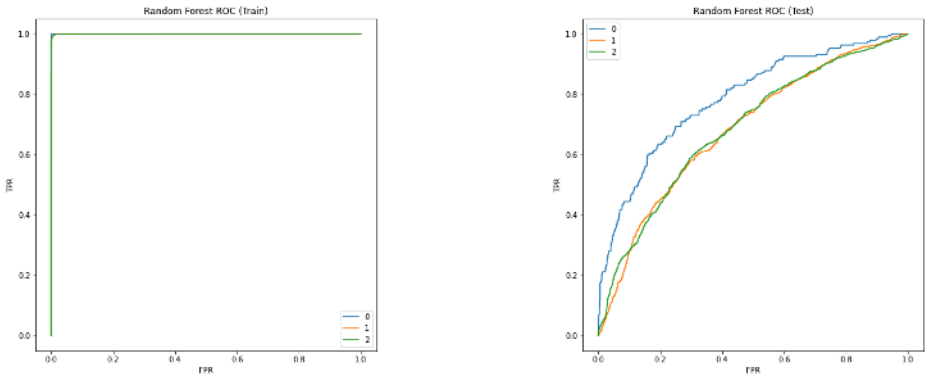
RESULTS

RESULTS – EXAMPLE



Confusion matrix for the best neural network model, showing that each class has been correctly predicted its own class more than any other class. Numbers are in percentages.

RESULTS – EXAMPLE



Example ROC curves for the random forest model on train (left) and test (right) data. Dramatic decrease going from train to test suggests overfitting.

CONCLUSIONS

- Inference: investigating tweet data reveals psychology behind positive and negative reviews
- Predictive power is decent, but better model performance can likely be achieved
- Productionization to expand project's impact

FUTURE WORK

- Improve model performance
- Further text feature engineering
  - Bigrams
  - POS tagging
- Productionization

ACKNOWLEDGEMENTS

Thank you!

Special thanks to Jacob Eli Thomas and Victor Geislinger, my instructors at Flatiron School, for their encouragement, instruction, and guidance.