

What's in a Tweet?



Phase 4 Project Presentation
By Ashley Eakland and Jose Harper

Agenda

Business Problem

Data Highlights

Modeling

Final Results

Business Problem

**Ensure best possible
experience for consumers**

- Accurately predict consumer sentiment based on tweet content
- Surrounding specific companies and products

— — —

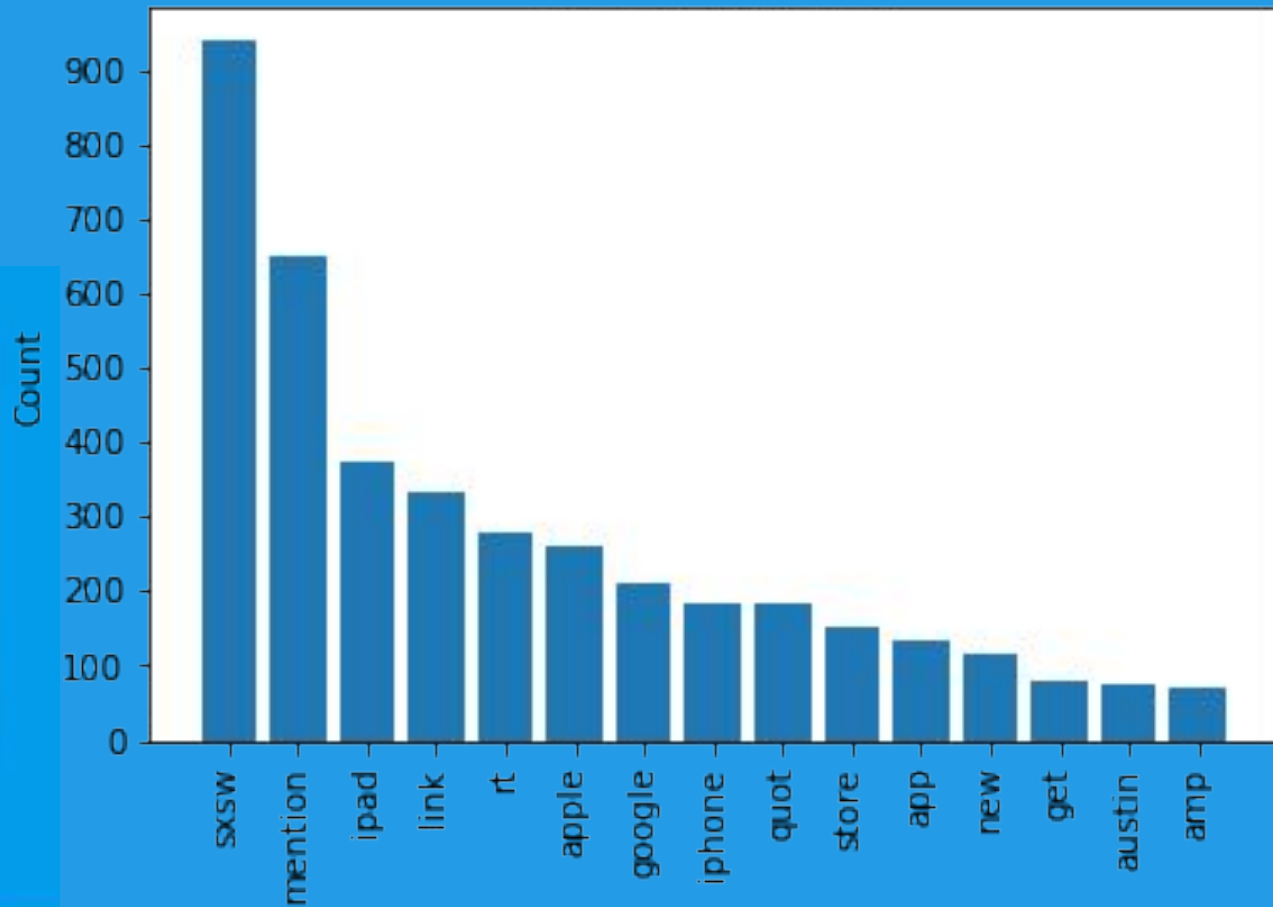
- ◆ 9093 tweets from CrowdFlower, supplied by data.world
- ◆ Human labeled
 - ◆ Emotions towards company/product mentioned
- ◆ Binary classification
 - ◆ Neutral reviews for future analysis

Data Highlights

Data Highlights cont.



Frequency Top 15



Data Preparation

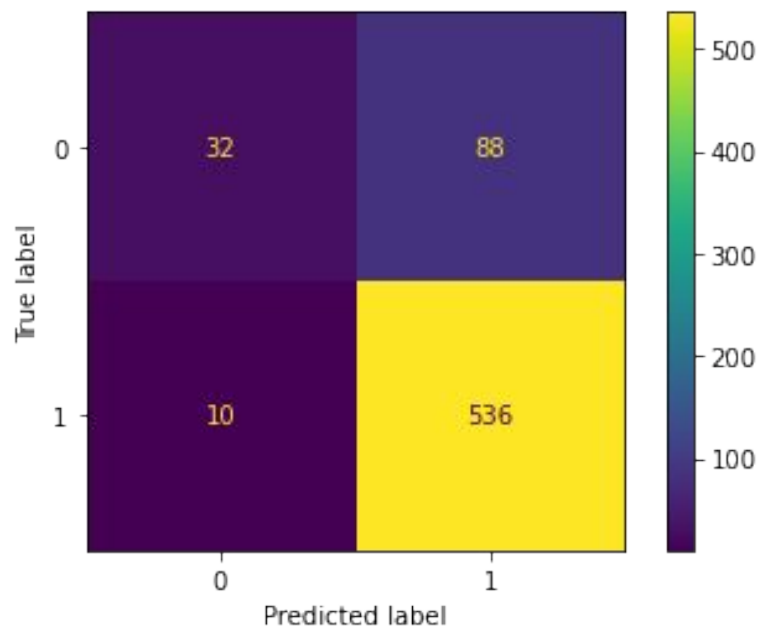
Train and Tune

Retrain and Validate

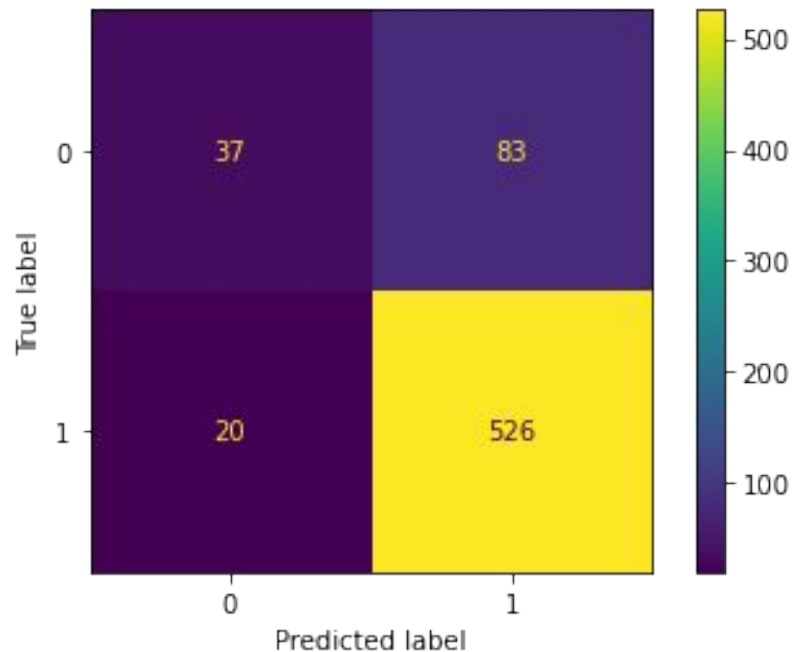
Final Evaluation

Modeling Process

Modeling Process cont.



RandomForest Validation



MultiNomial Bayes Validation



Final Results

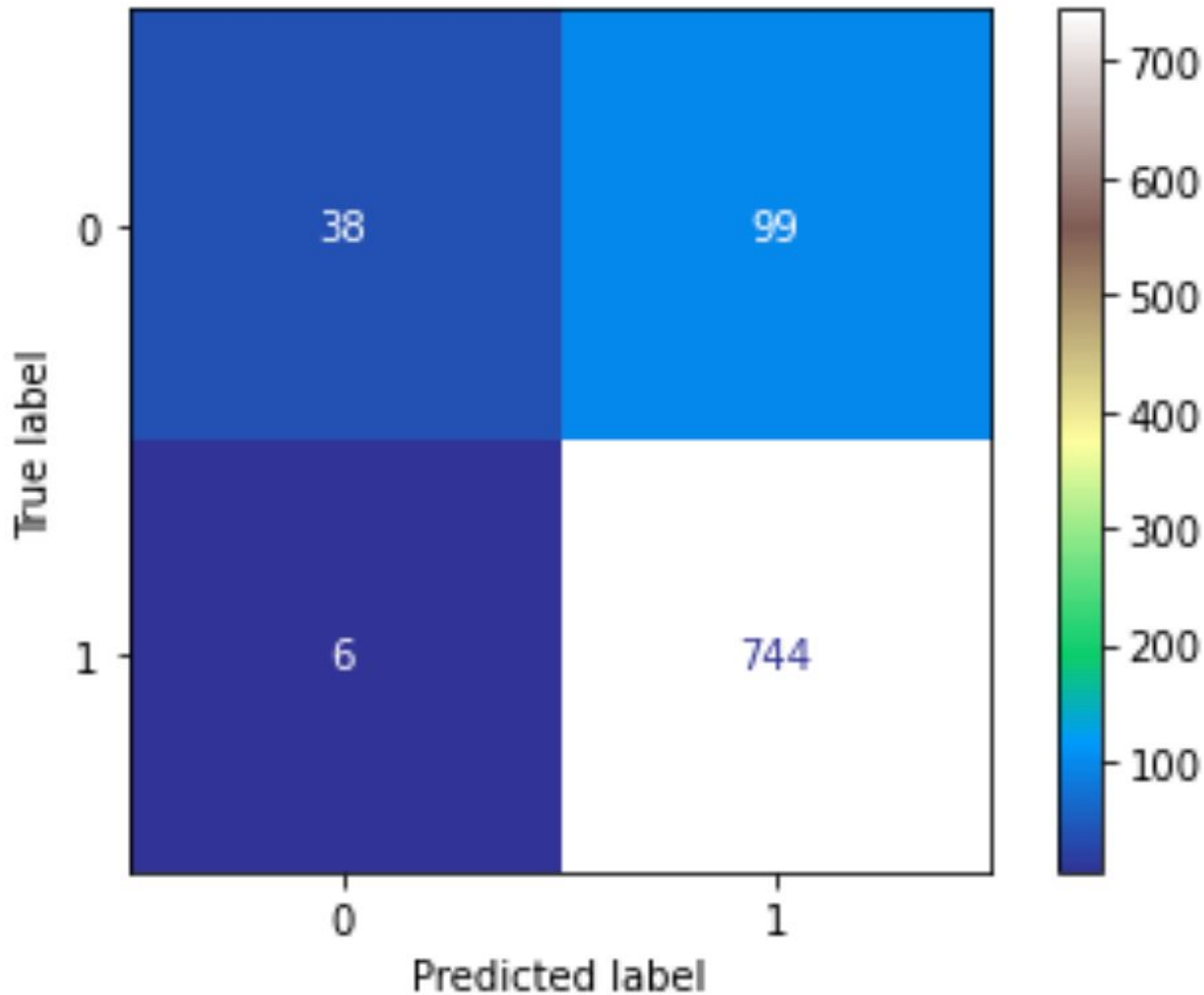


Final Results

88% Accuracy



88% Precision



RandomForest Classifier

TF-IDF Vectorized Data

4150-Word Vocab

Fail

Headaches

Money

iPad

Key Features

Key Features cont.

Fascist

Apple

Sucks

Hate

*Long, Fades**



Future Analysis



- ◆ Add in Neutral
- ◆ Deep dive into hashtags/Tweet specific acronyms
- ◆ Product/service is associated with positive/negative



Questions?

Questions?

Thank you!

Ashley Eakland | ashley@eakland.net | LinkedIn <https://www.linkedin.com/in/ashleyeakland>
Jose Harper | jose.harper@gmail.com | LinkedIn <https://www.linkedin.com/in/jose-harper-36a95489/>

For technical information and to see the Jupyter Notebook:
GitHub Repository / www.github.com/smashley-eakland/tweet-nlp-analysis