Determining Product Sentiment Using Natural Language Processing (NLP)

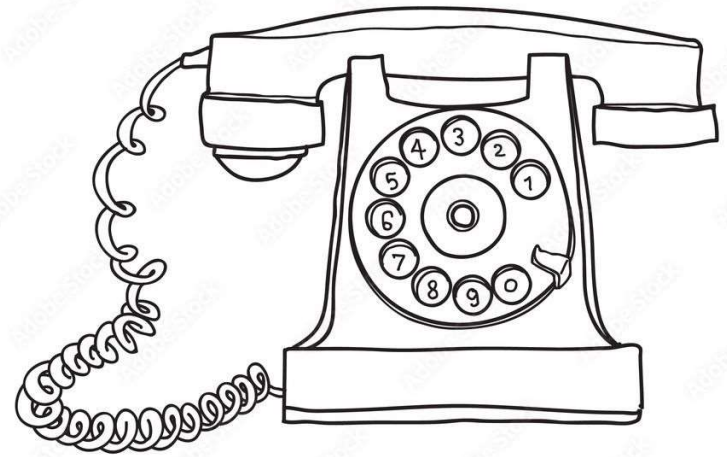
Matt LeRoi
Sep 18, 2025

Business Problem

A stealth tech company wants to create a fancy new device. They want to:

- Flag positive tweets about existing products
- Maximize number of tweets available for further analysis
- Minimize tweets mislabeled as positive

- The F1 score is a good metric for balancing these goals

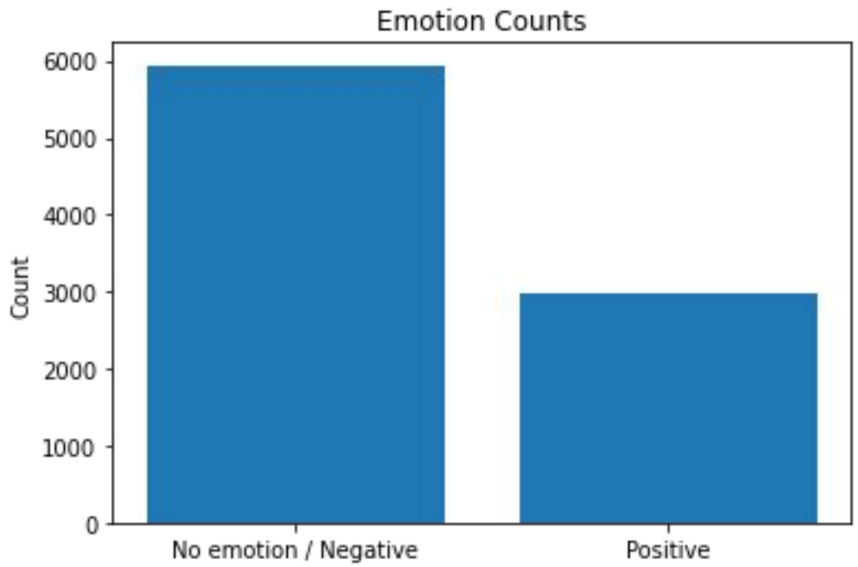


Data

There are ~9000 tweets total

After combining negative and neutral tweets:

- Negative/Neutral: 6,000
- Positive emotion: 3,000



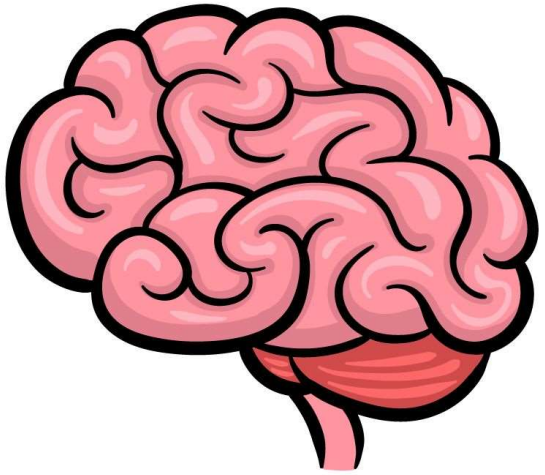
Data limitations (aka: emotions are tricky)

- All data is labeled by humans, using human judgement
- Data is outdated (from 2013)

Two Models

Created:

- Logistic regression model
- Sequential neural network model



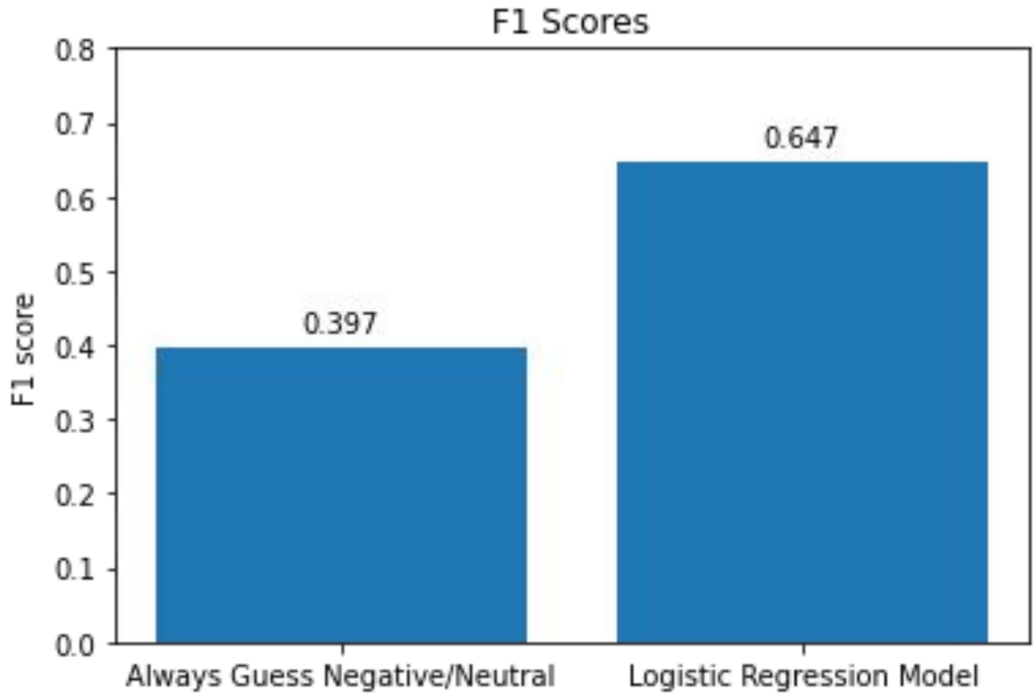
Baseline model result

Simplest model:

- Always assume the majority class (negative/neutral)
- Low F1 score due to 0% of positive tweets identified
- F1 score: 0.397

Logistic regression model:

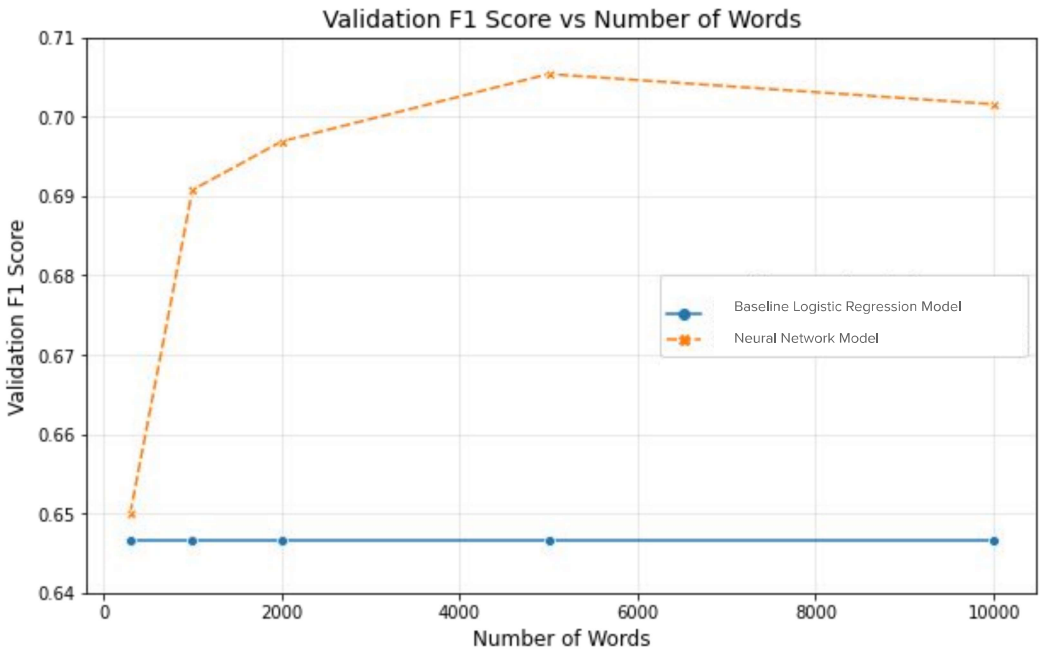
- F1 score: 0.647



Neural network model validation result

- Number of words:
  - How many words are included in the model

- Max F1 score: 0.705
  - 5000 words



Most positive words:

Baseline model:

- 'cool'
- 'great'
- 'ipad'
- 'good'
- 'love'
- 'wow'
- 'awesome'
- 'nice'
- 'excited'
- 'fun'

Neural network model:

- 'cool'
- 'excited'
- 'brilliant'
- 'congrats'
- 'wow'
- 'great'
- 'smart'
- 'genius'
- 'woot'
- 'zomg'

Final test result

Winning configuration:

- F1 score: 0.667
  - Similar to validation result
  - An F1 score of 0.7 is generally considered acceptable to moderately good

Next steps

- Further tuning and other model types
- Verify ground truth by reclassifying tweets with multiple humans to ensure greater consistency in the labeled training data
- Analyze full list of most positively associated words

Thank You!

Email: mcleroi@gmail.com
GitHub: @mattleroi
LinkedIn: linkedin.com/in/mcleroi/