Apple Twitter Sentiment Analysis

A PROTOTYPE RNN

Business Problem

Traditional market analysis is:

- Expensive
- Time Consuming
- Low in Information Content

A NLP model can allow for more efficient processing of large amounts of available data on public sentiment on social media

Project Goals

- Determine what kind of model would be best of this task
- Test the effects of various model features
- Construct a prototype sentiment analysis model

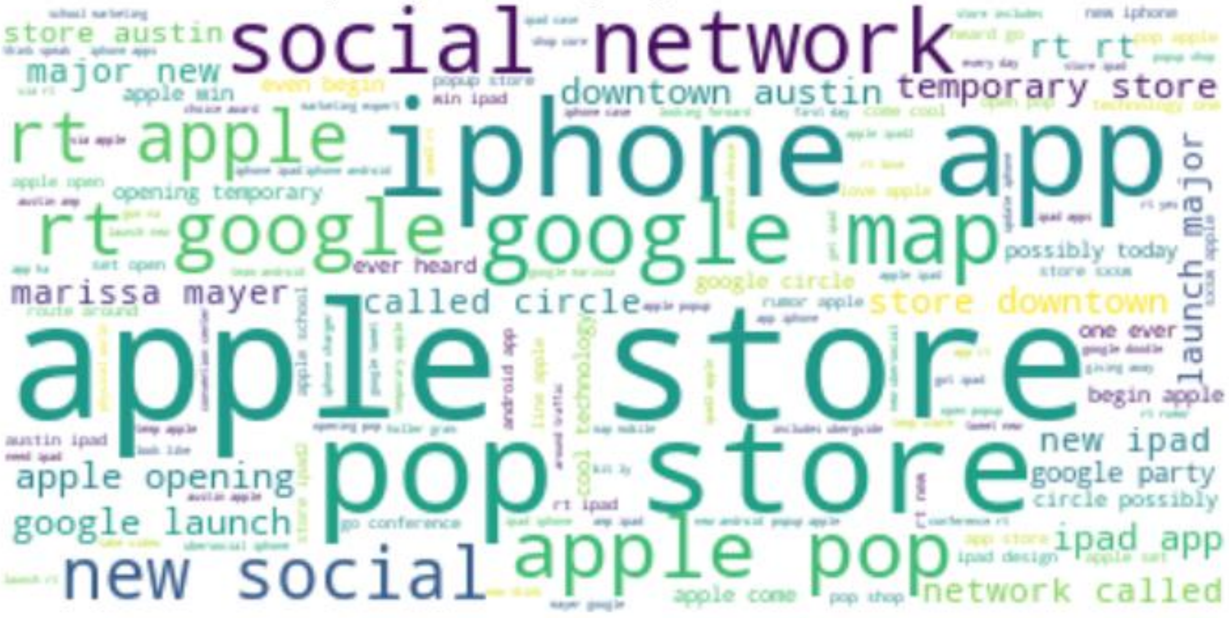
Data

Data was taken from data.world and consists of:

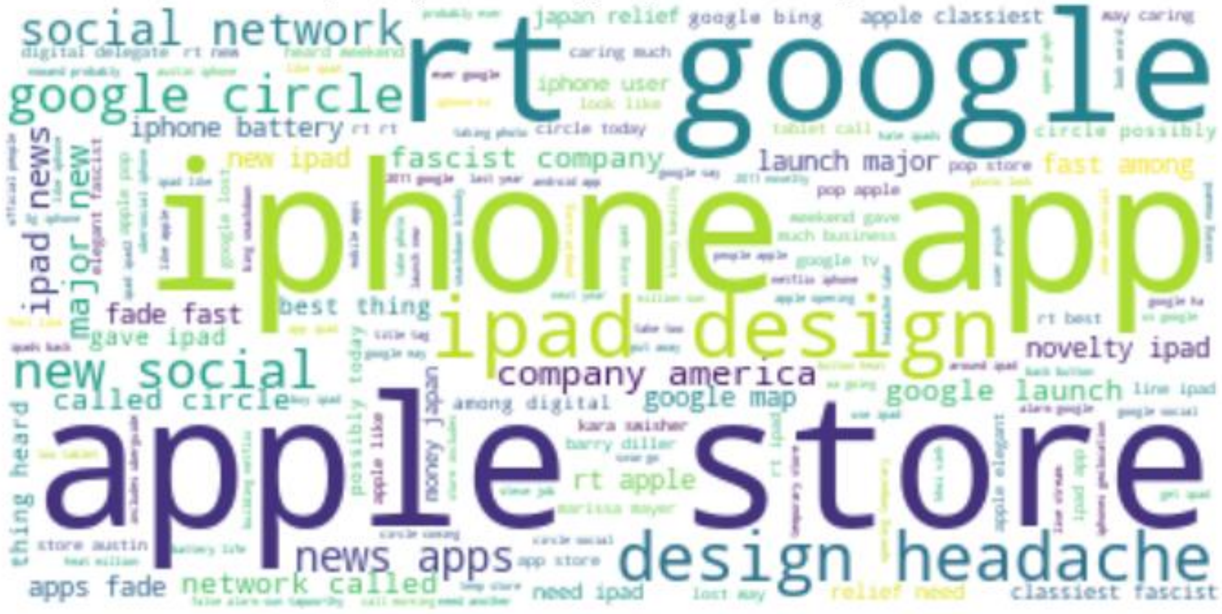
- 9203 tweets from the #SXSW hashtag
- All tweets contain keywords mentioning Apple or Google products
- Tweets were labeled by crowd sourcing as having either positive, negative or neutral sentiment to the brand
- 61% of the tweets were neutral, 33% positive, and only 6% negative

Visualizing the Data

Most Frequently Occurring Bigrams in Positive Tweets



Most Frequently Occurring Bigrams in Negative Tweets



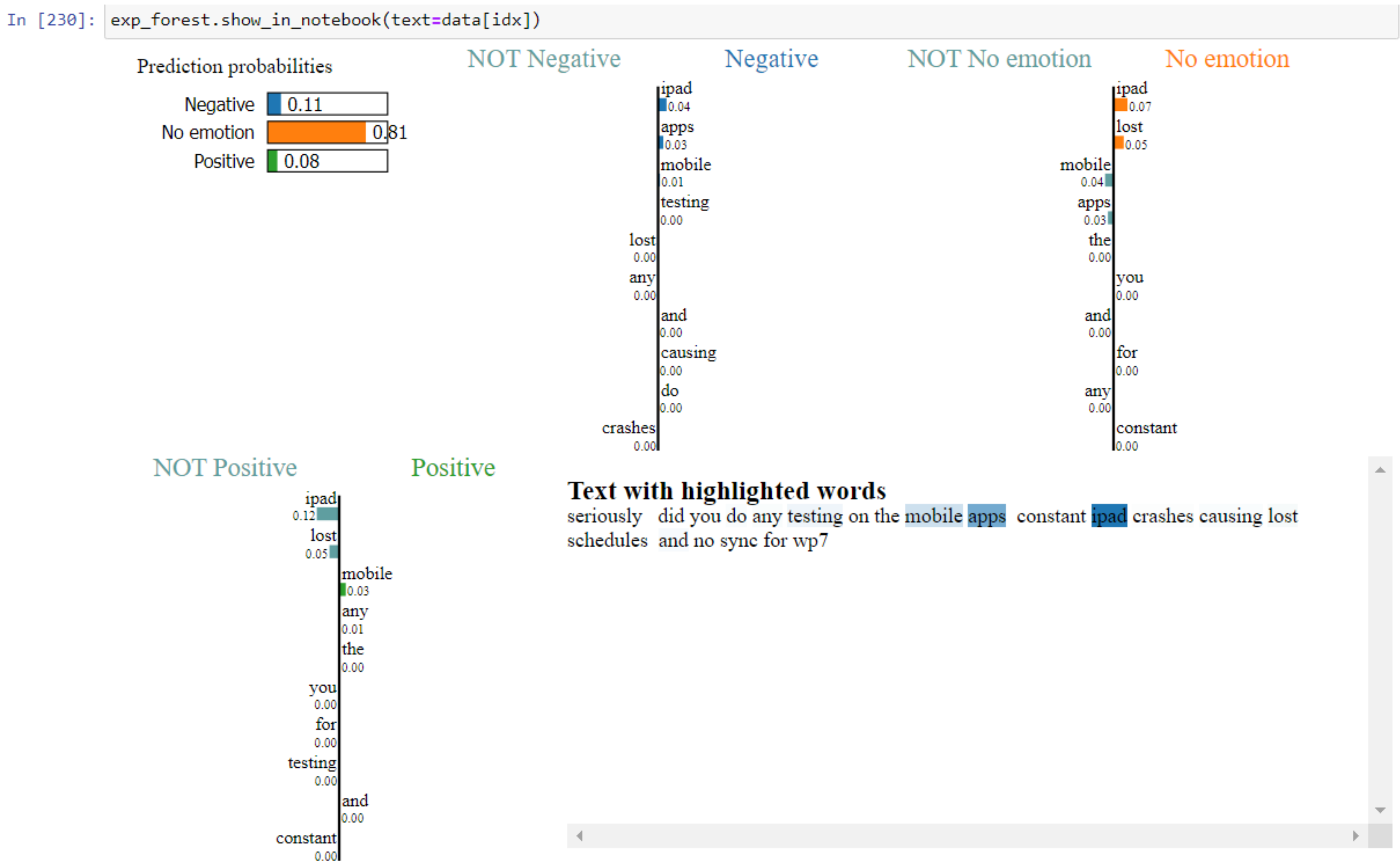
Models and Features Tested

Models

- Naïve Bayesian
- Logistic
- Random Forest
- SVC
- RNN

Features

- GloVe vs Custom embedding
- Class Weighting
- Augmentation with Synonyms
- Depth
- Number of Nodes

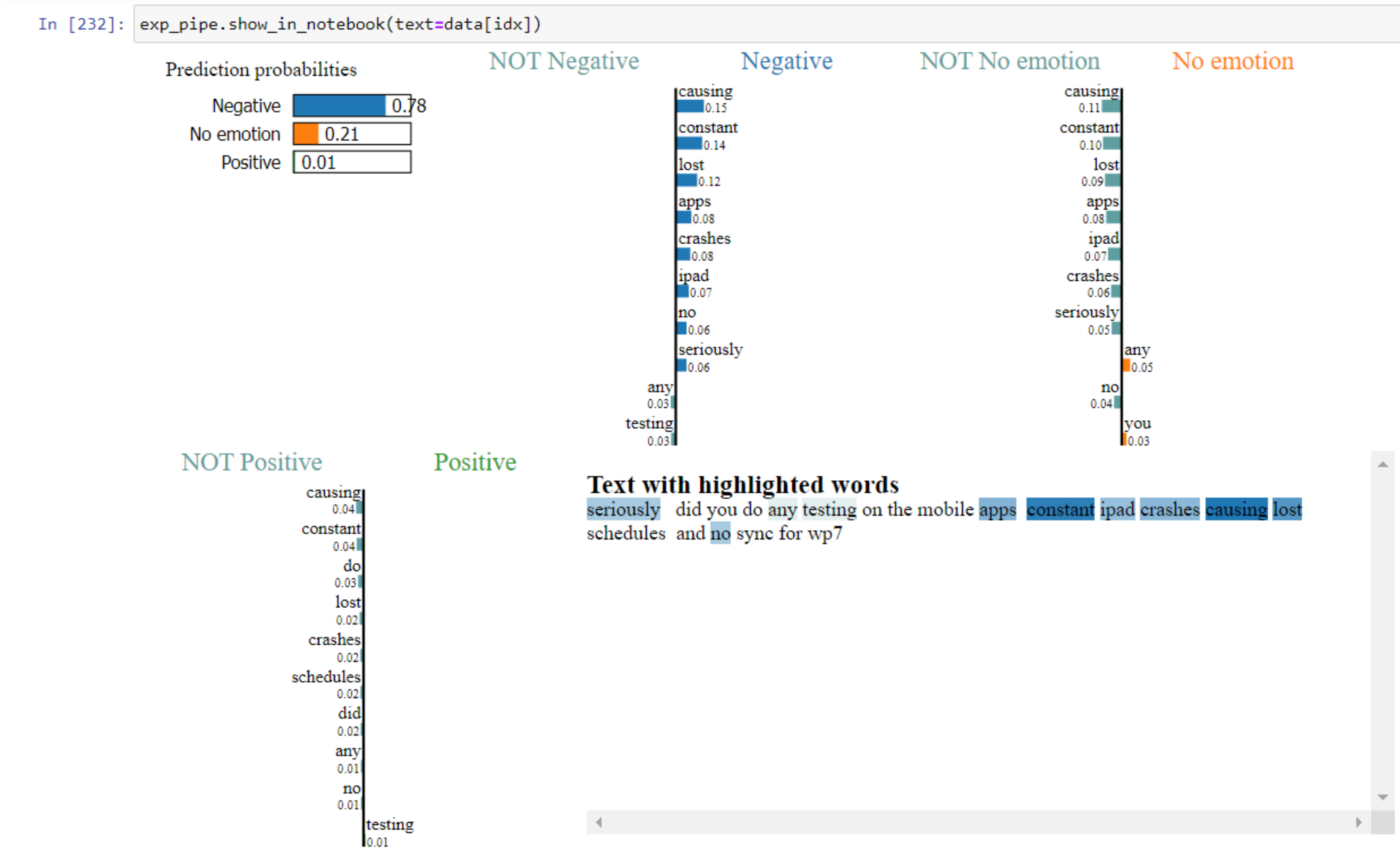


Limits of Tradition Models in NLP

We can see in this example that:

1) The model failed to recognize this was a negative tweet

2) It failed because it was unable to recognize the relevant features



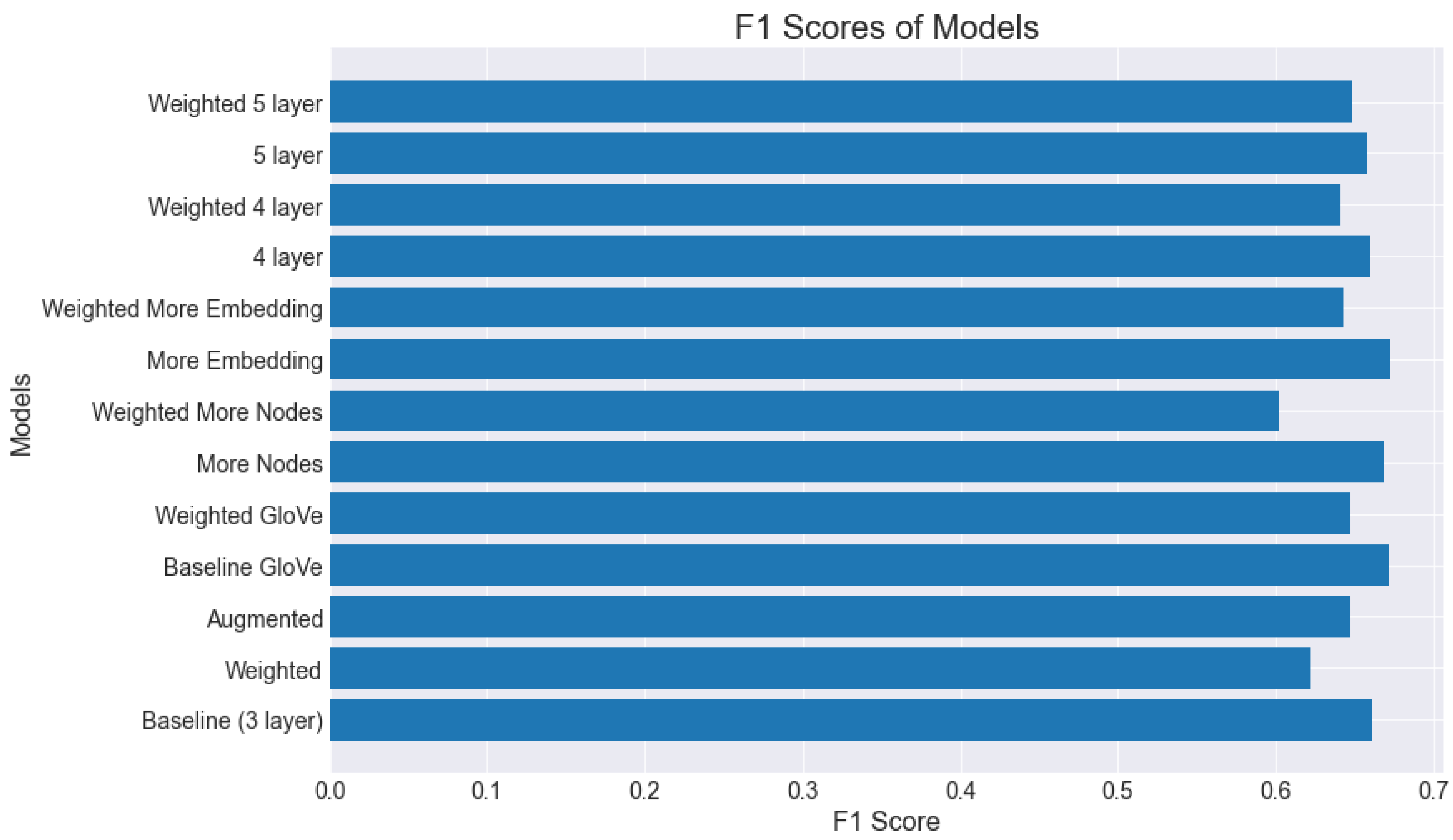
Advantage of RNN in NLP

We can see in this example that:

1) The RNN correctly identified this was a negative tweet

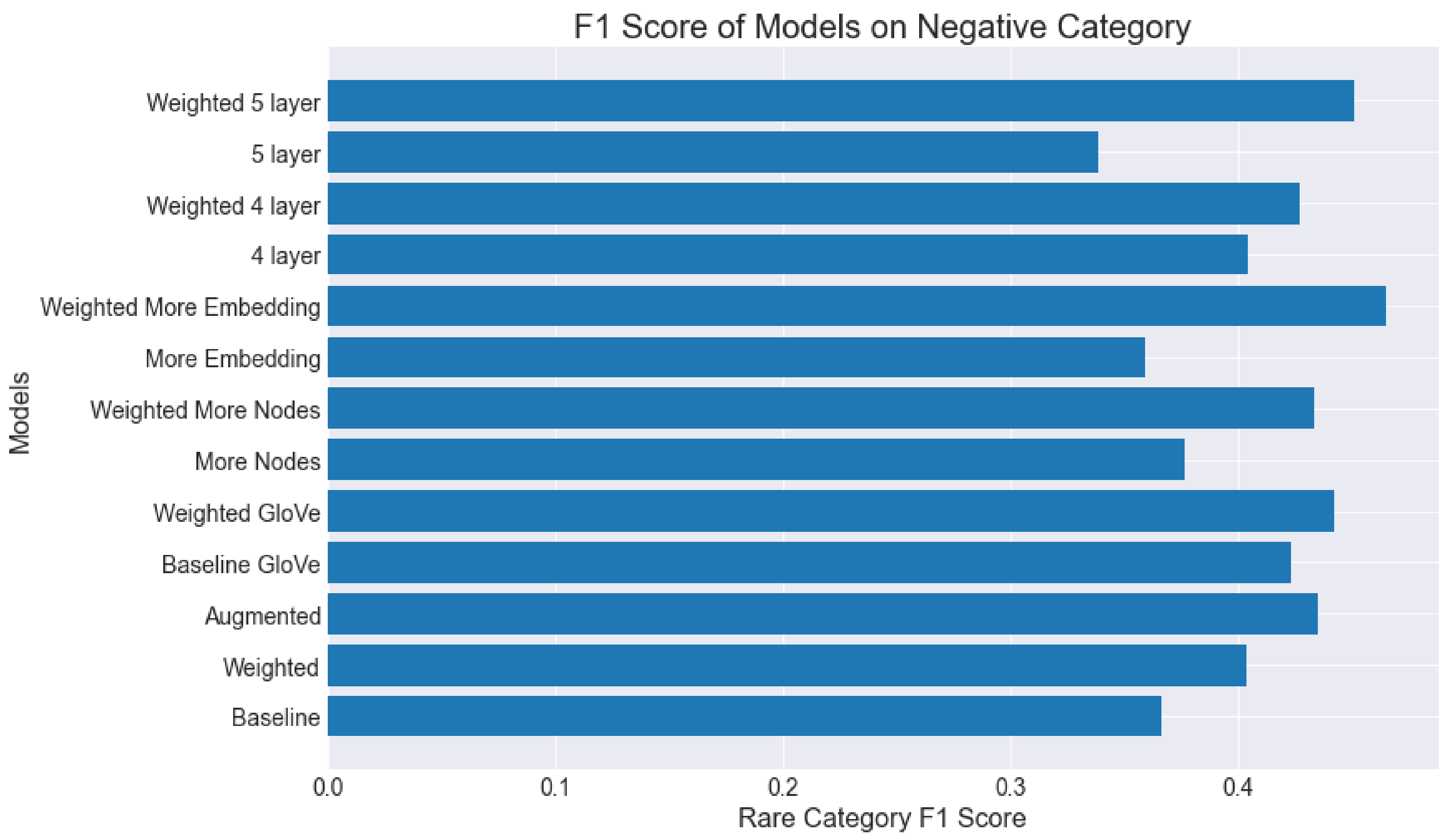
2) It correctly picked up on some of the clearly negative wording, though not perfectly

Comparing Model Features by Overall Performance



- Weighted models tended to do worse overall
- GloVe embedding showed some improvement
- Synonym augmentation did not improve performance
- Differences in structure were minimal on this data set

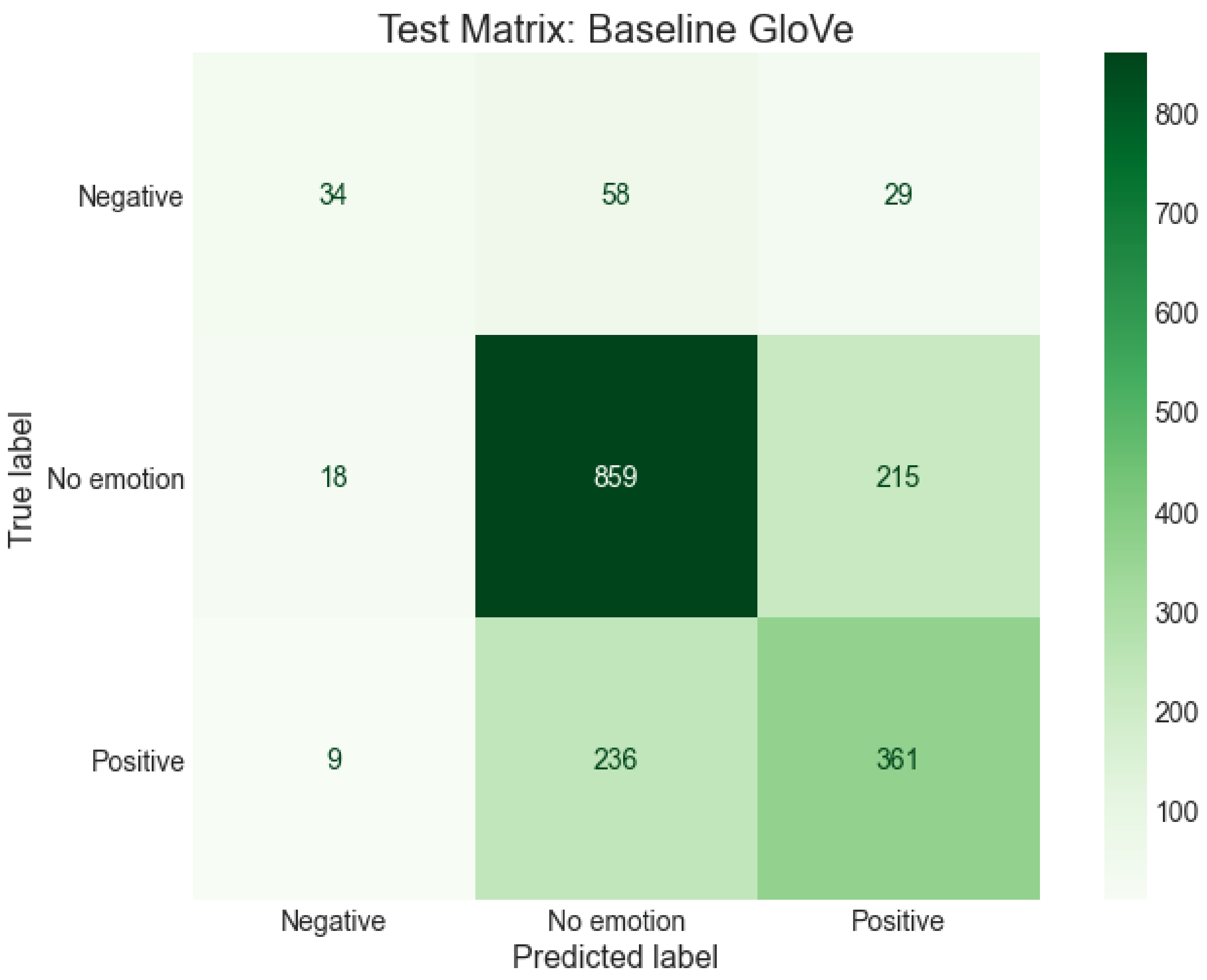
Comparing Model Features by Performance on Rare Data



- Weighted models tended to do better on rare data
- Pretrained or more embedding improved performance on rare data
- Deeper networks were somewhat better on rare data

Final Model On Test Data

GloVe model had the highest weighted F1 score of 68%



Conclusions

- RNN are promising tools for sentiment analysis and should be used in this kind of analysis *given sufficient data*

- Improvements by using RNNs can be small if the data is limited, there was only a 4% improvement in F1 score (68% for RNNs versus 64%) using a dataset this small

Limitations

- Data was specifically taken from a hashtag that was more likely to contain tweets talking about brands and products. This potentially places limits on its generalizability

- Given the extreme class imbalance in this data set, the performance on the rare category is not very likely to be reliable

Contact Information

Name

Lia Elwonger

Email

lelwonger@gmail.com

Github for this project

https://github.com/nonlocal-lia/sentiment-analysis-project