Twitter Sentiment Analysis:

Target Retail Store

Henry Van Gorp

Outline

- Business Problem
- Data Understanding
- Natural Language Processing
- Modeling Process
- Conclusion and Recommendation

BUSINESS PROBLEM

- Target is one of the world largest retailers with a big electronic department in most stores.
- As the holiday season is approaching large retail sellers like Target are beginning to focus on what should be showcased in their stores especially during times where sales occur.
- Target is looking to better understand the sentiment behind the products which they are selling in their electronics department.
- By building a model that can use twitter data from one of the worlds largest technology conferences they will better understand what is looked upon positively and what is looked upon negatively from potential buyers.

DATA UNDERSTANDING

- The dataset comes from CrowdFlower via data.world.
- Human Raters rated the sentiment in over 9,000 Tweets as positive, negative, or neither (neutral).
- The dataset shows that this twitter set was most likely pulled from those who were attending the SXSW conference due to the amount of mentions in the tweets.

POSITIVE WORD CLOUD

NEGATIVE WORD CLOUD

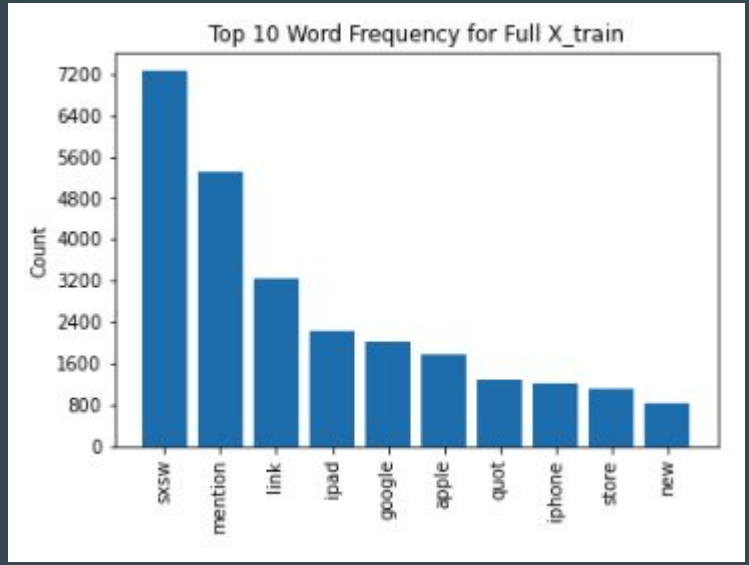
NEUTRAL WORD CLOUD

NATURAL LANGUAGE PROCESSING

Natural Language Processing was utilized to clean the twitter data. This process included:

- Standardizing
- Tokenizing

After cleaning the text data, you could look at frequency of words in the full dataset.



MODELING

Three different models were created before choosing the best fit model. Before modeling was done, the data was vectorized using TF-IDF Vectorizer.

Model Types:

- Baseline Model with Multinomial NB
- Random Forest
- XGBoost

CONCLUSION

- The best fit model was built with Random Forest.
- This model will accurately predict the emotion of a tweet 67% of the time.
- There was a huge class imbalance causing Class 0 (negative emotion) to perform poorly. Class 1 (neutral emotion) performed the best in the final model.
- It is recommended to utilize this model to predict the emotion behind specific products which will be sold at Target.
- It is also recommended to utilize this model with more data to increase the scores.