Phase 4 NLP Project

Mark Rubin

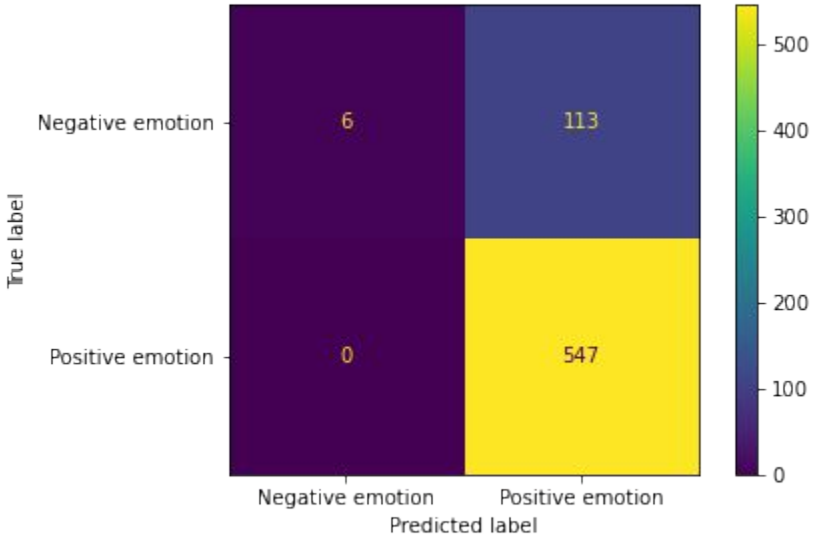
Business Problem / Data

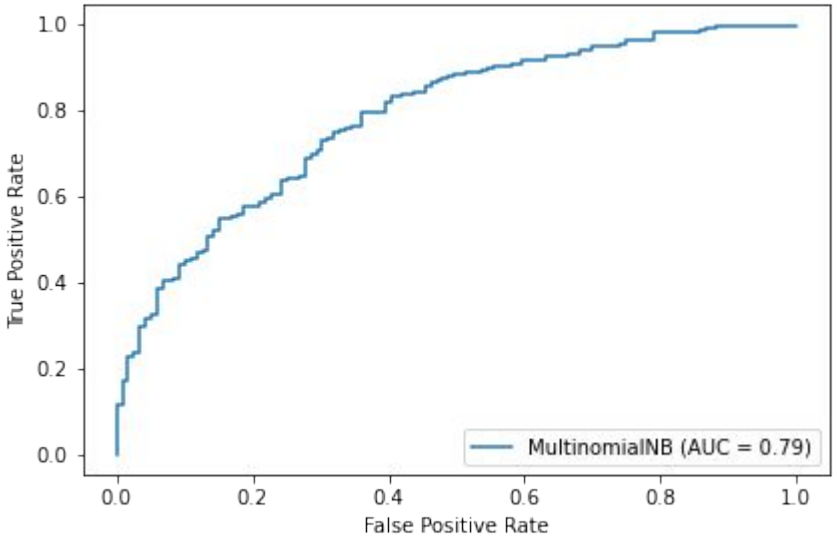
- **Business Problem:** Major brands would like to know how consumers feel about their various products.
  - Analyze Twitter user's sentiments about Apple and Google products to better understand how consumers feel about their products.

- **Data:** 'Brands and Product Emotions'
  - Dataset found on CrowdFlower via data.world
  - Human raters rated the sentiment in over 9,000 Tweets as positive, negative, or neither.
  - Class Imbalance:
    - 2978 positive tweets
    - 570 negative tweets

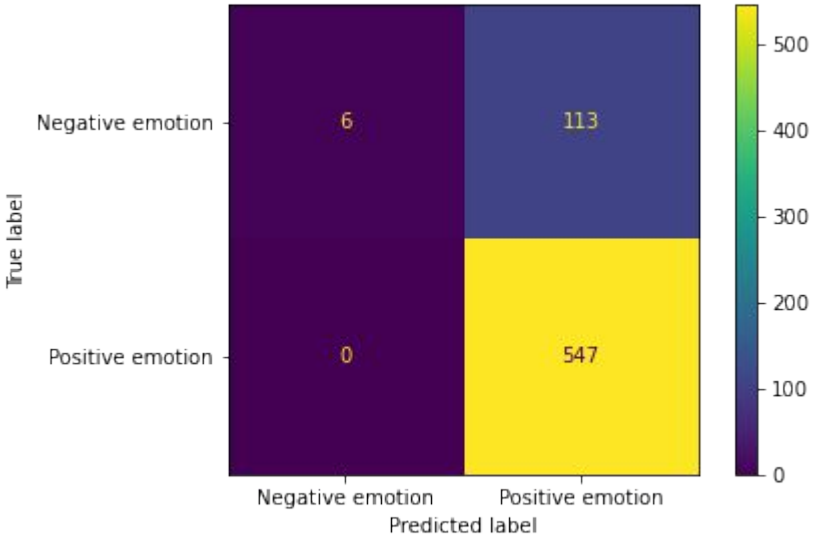
Modeling Process:

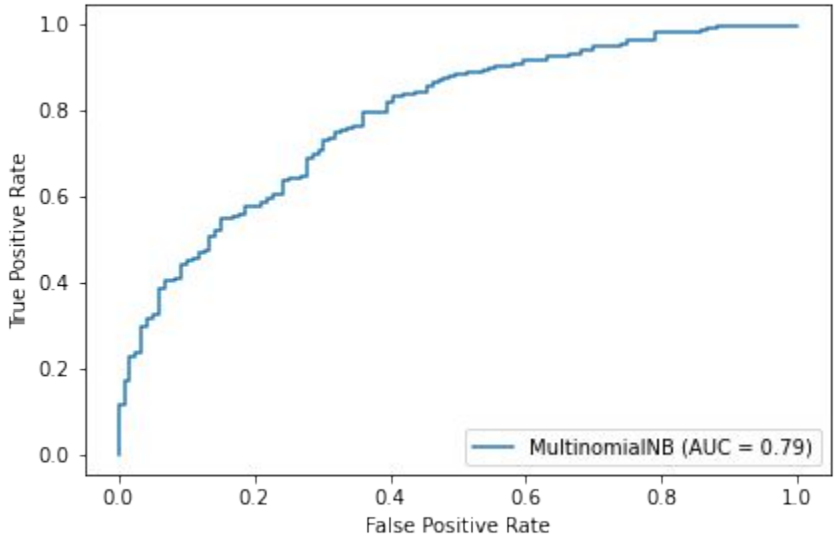
Data processed using a Count-Vectorizer and Classified with MNB



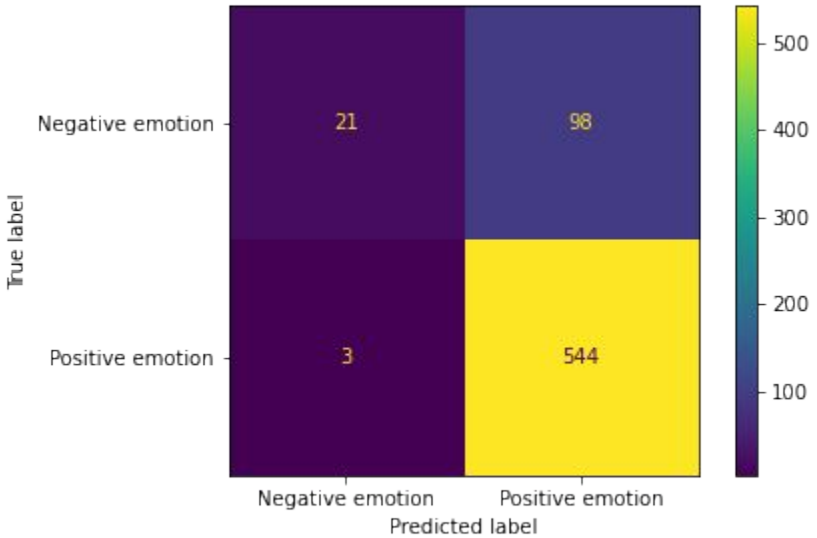


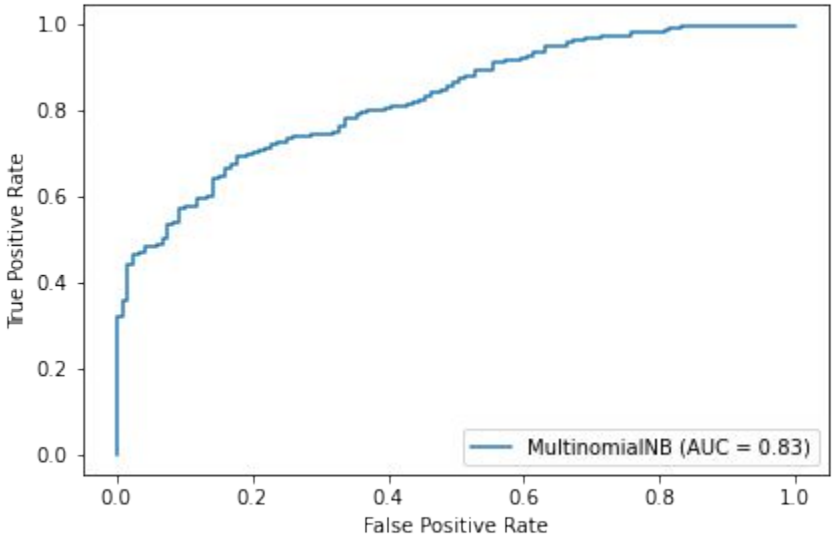
TFIDF Vectorizer and MNB Classifier



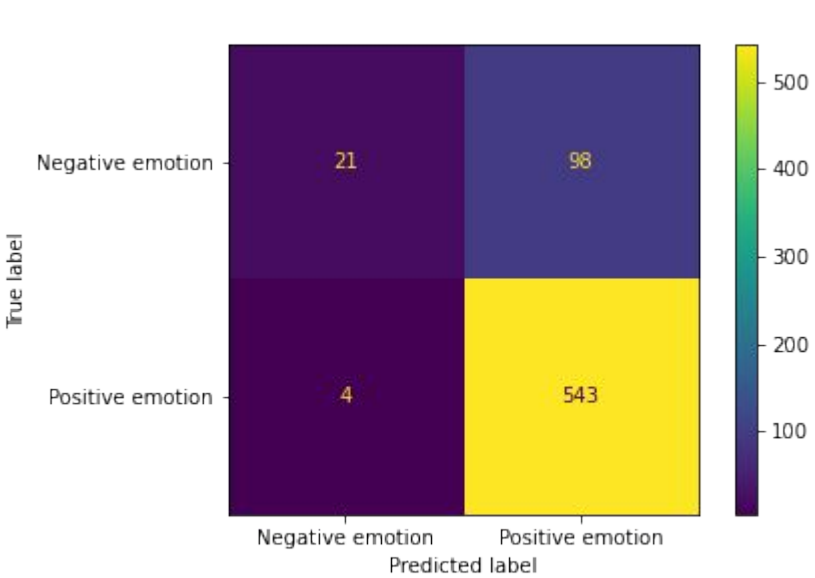


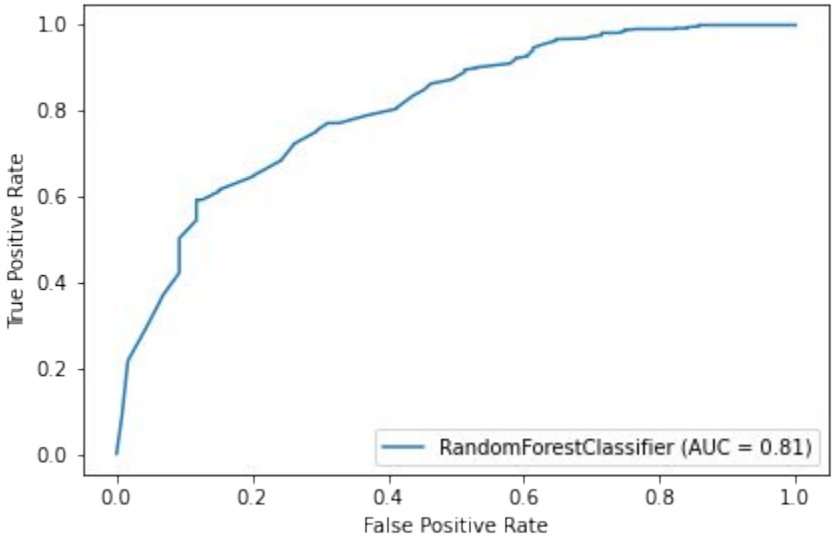
TFIDF Vectorizer and MNB Classifier (Alpha=.2)



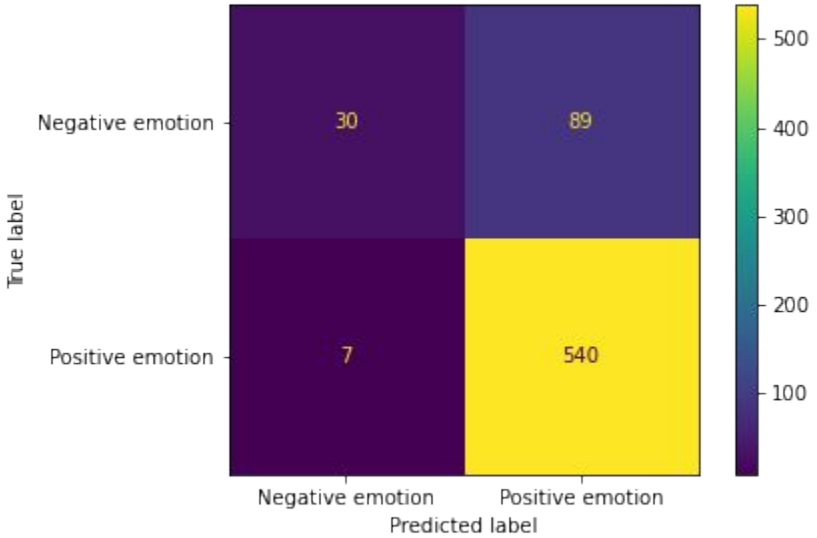


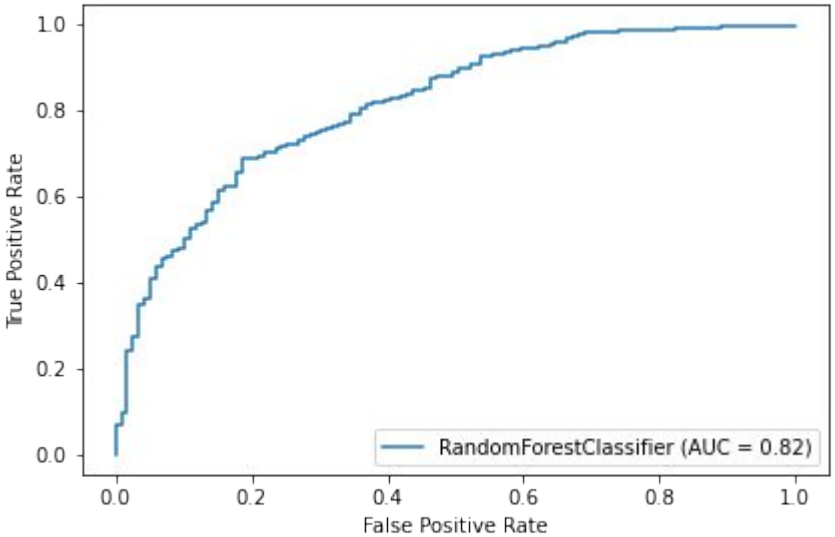
Random Forest





Random Forest w/ Weighted Classes





Results and Further Steps.

- Best Model Achieved: Random Forest w/ Weighted Classes
- Model Performed Decently well
  - Precision Metric Score of 86%.
  - Accuracy Metric Score of 85%.
- Moving forward:
  - I would like to include a Neutral Class to change this to a multiclass classifier.