



Sentiment Classification of Album Reviews

By Daniel Mocombe



Business Case

- Metrics for success in the music industry continue to evolve
- Digital album sales dropped 7.3% between 2017 - 2018
- Streaming makes it difficult to measure overall album performance
- Direct feedback gained by online user album reviews
- Review sentiment helps artists and record labels understand fans' likes/dislikes



The Data

- Album reviews from users on metacritic.com
- Classified into positive, neutral, and negative reviews
- Analyze album reviews
- Look into the words used in different review classes
- Create an NLP classification model



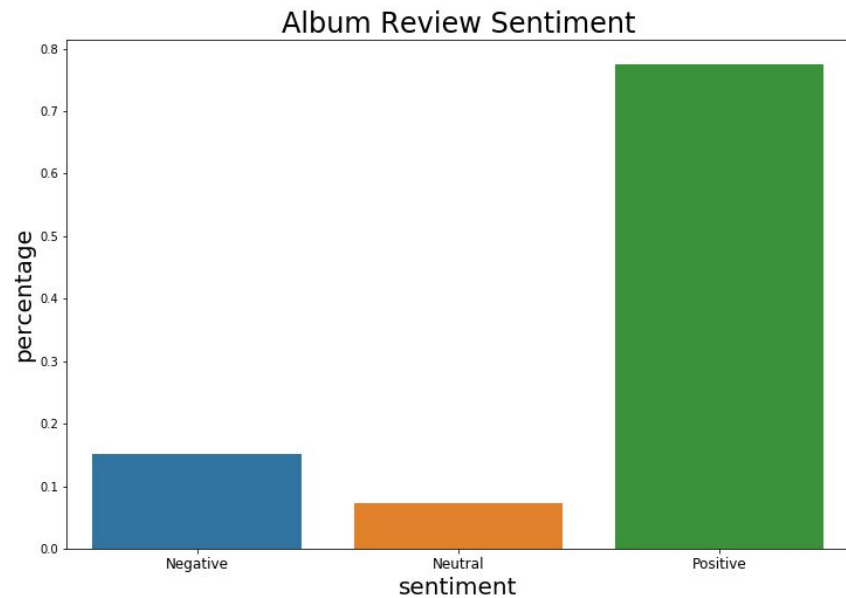
Models

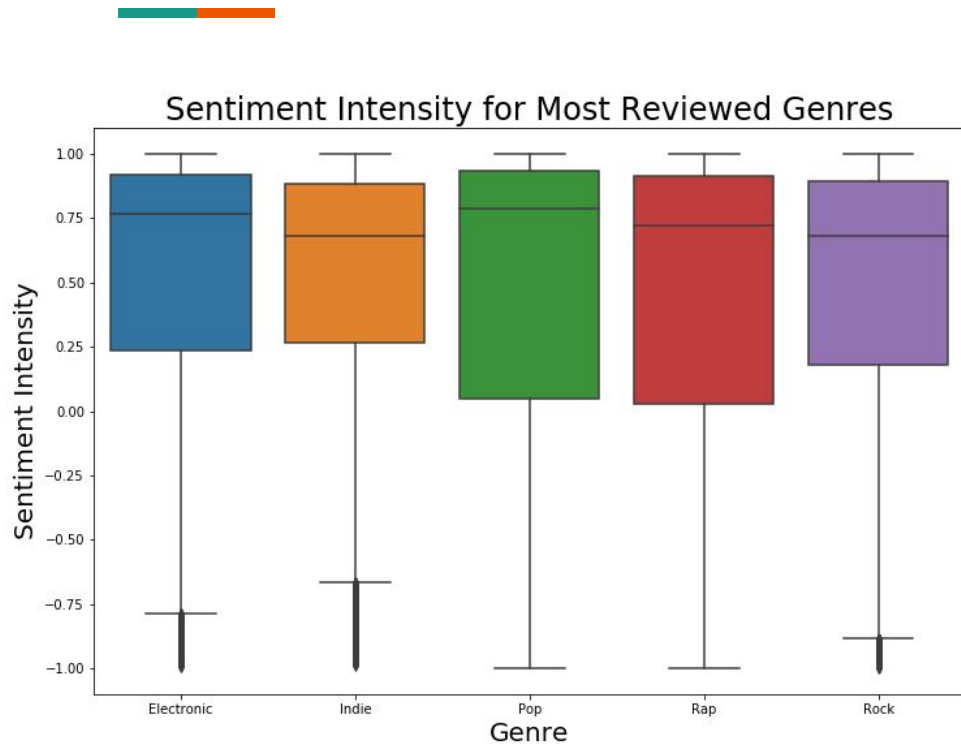
Model	Accuracy	F1 (Weighted)
(Baseline) Dummy Classifier	0.778	0.680
Random Forest	0.848	0.819
Naive Bayes	0.789	0.718
Support Vector Machine	0.887	0.877
(spaCy) Random Forest	0.844	0.812
(spaCy) Naive Bayes	0.790	0.720
(spaCy) Support Vector Machine	0.778	0.875



Classes

- 84,054 user reviews
- Vader's Sentiment Intensity Analyz
- Classes highly imbalanced

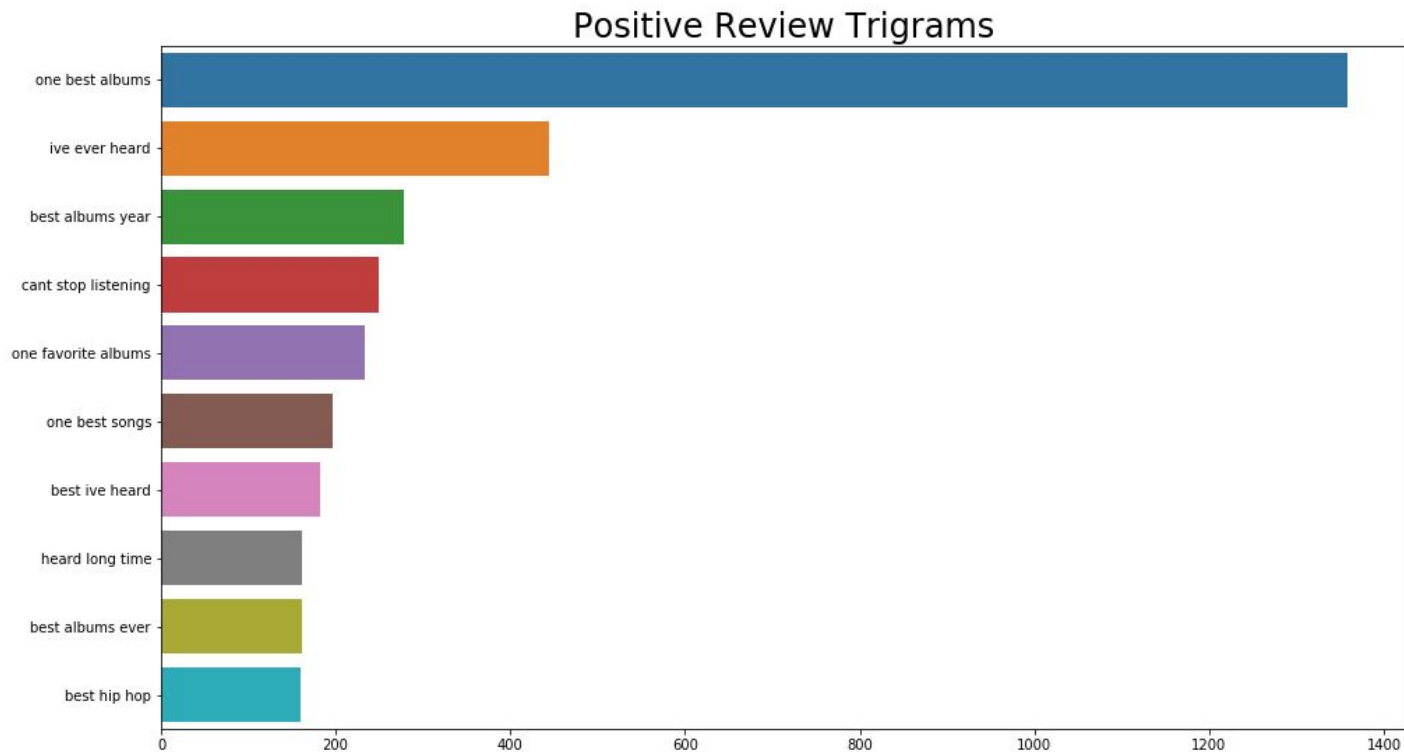




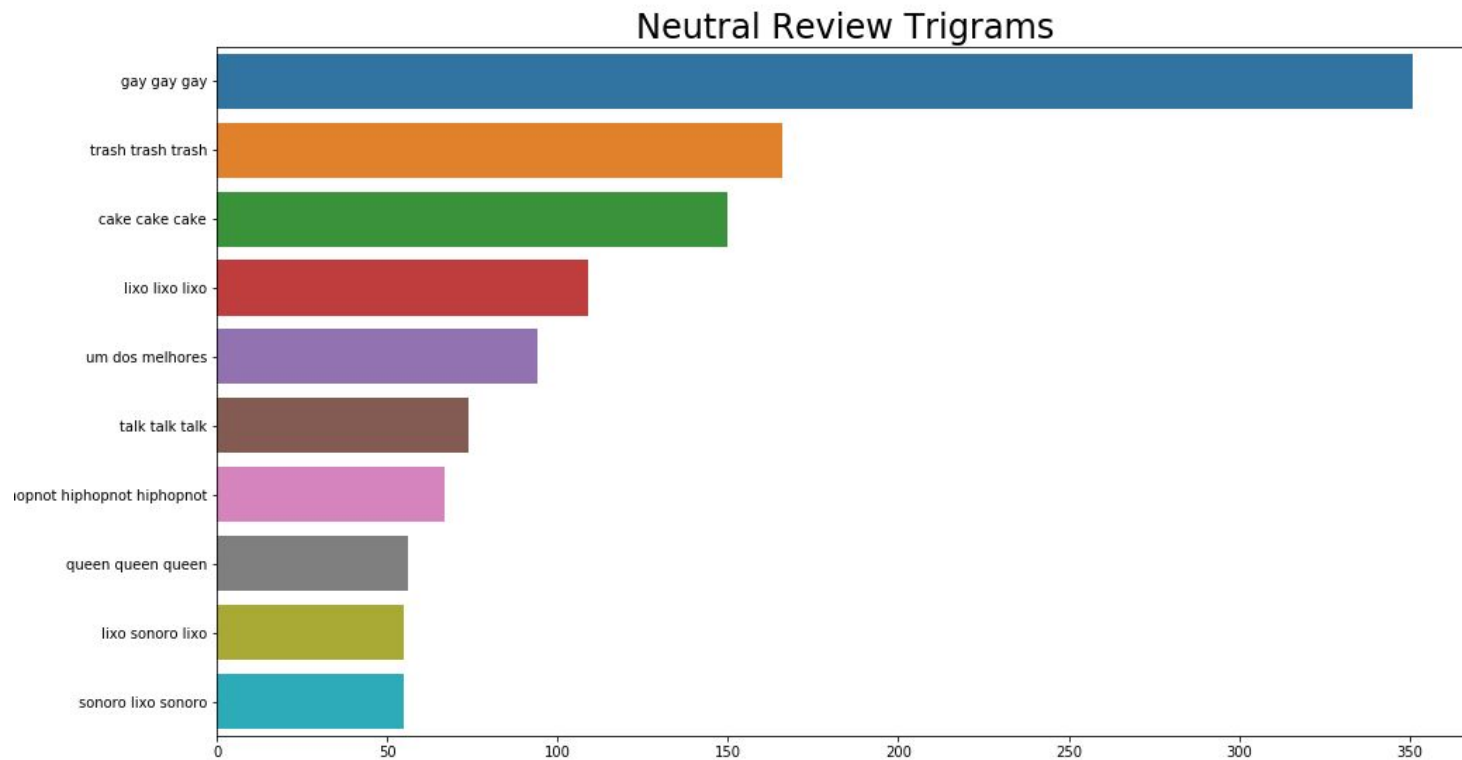
Genre Sentiment

- Positive - ≥ 0.05
- Neutral - < 0.05 & > -0.05
- Negative - ≤ -0.05

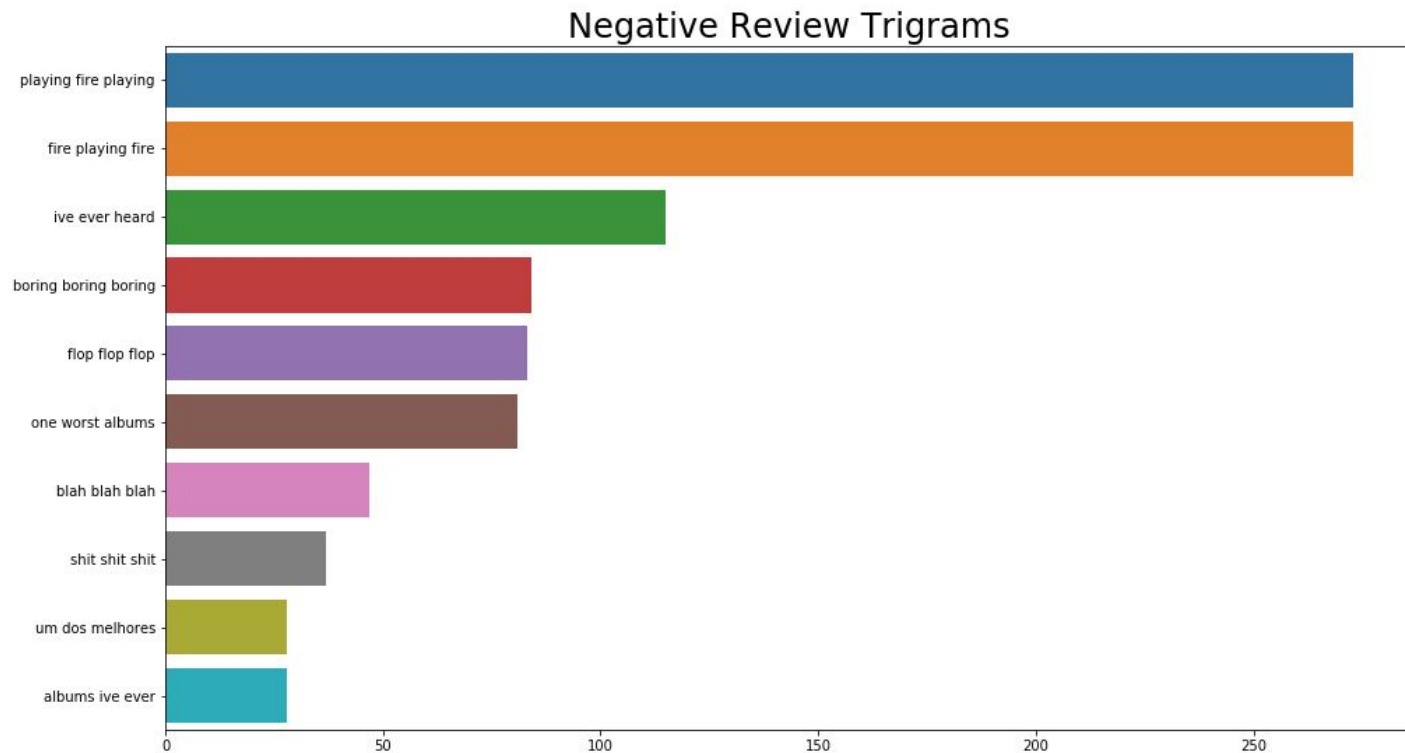
Trigrams



Trigrams

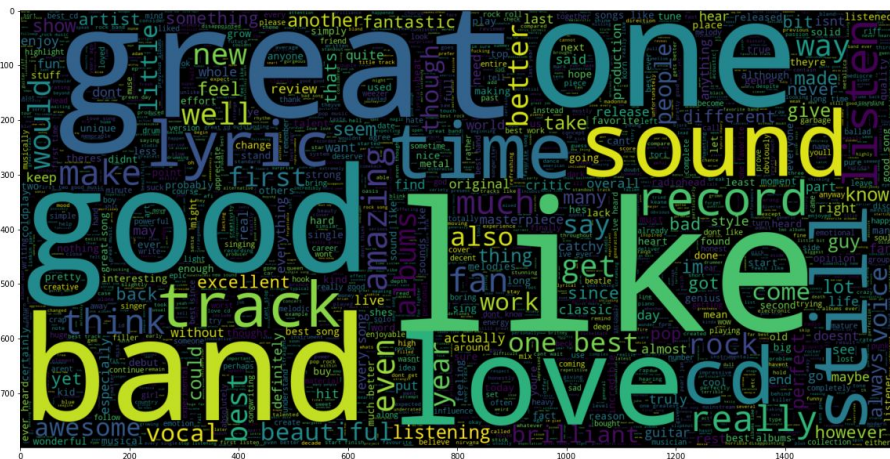


Trigrams

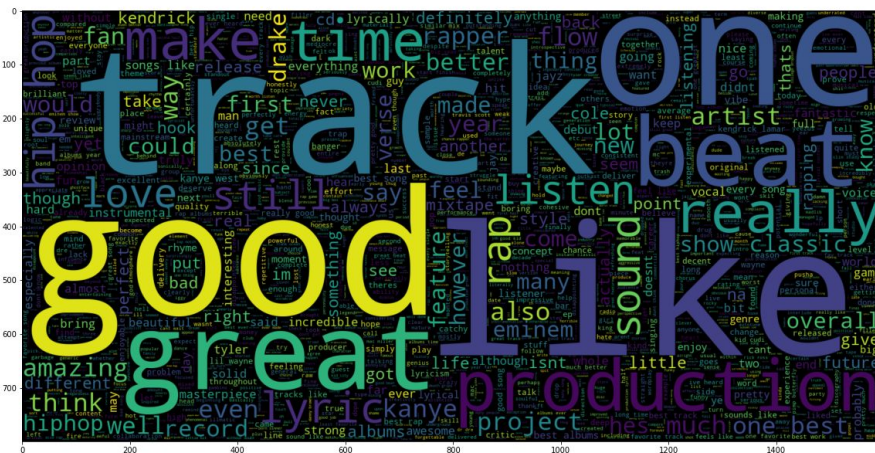


Positive Word Clouds for Rock vs Rap

Rock



Rap





Preprocessing and Modeling

- Updated stopwords list (album, music, song).
- Lemmatized review text
- TF-IDF used for word vectorization
- Random Forest, Naive Bayes, and SVM classifiers used for modeling
- spaCy used for modeling process as well
- Final Model parameters
 - SVM
 - Kernal = 'linear'
 - C = 1.0



Limitations and Discussion

- Some reviews misclassified in data preparation
- Irrelevant or misleading user reviews
- Rock fans mention lyrics and sound of the album
- Rap fans refer to specific tracks, and quality of the beats and production.
- Investigate differences in user rating vs sentiment classes
- Analyze critic reviews and topic modeling
- Review text generation.

THANK YOU