



Decoding Album Reviews for Pink Floyd's Dark Side of the Moon: Sentiment Analysis using NLP

By Nicholas Bronson



Agenda

1

Objectives

2

Methodology

3

Results

4

**Recommendations
& Next Steps**



1 Objectives:

My client is a record label that is seeking to analyze text data that they have acquired regarding artists on their label and their peers.

They aim to conduct a larger scale analytics initiative, however, they wish to better understand how large text-based datasets can be analyzed.

Key Questions:

- 1) Can text data be used to categorize and interpret album reviewer sentiments?
- 2) What are listener sentiments on Pink Floyd's Dark Side of the Moon, what are the perceived strengths and weaknesses of the album?

- **1547** Reviews
- Average of **138 words** per review
- **1494** reviews **contain a rating** (1-5)
- Average Rating is **~4.4**

- **NMF** is the primarily tool used

- Examination of **classification** methods including: **Logistic Regression, Random Forest, Gradient Boosted Algorithms**

Data Exploration

Pre - Processing

Dimensionality Reduction

Topic Modeling

Pre - Processing Round 2

Topic Modeling Round 2

Supervised Learning

Conclusions, Outline of Next Steps

2

- Removal of punctuation, upper case letters
- Tokenization

- Addition of **stop words** in **English** and **Spanish**
- Lemmatization

Methodology:



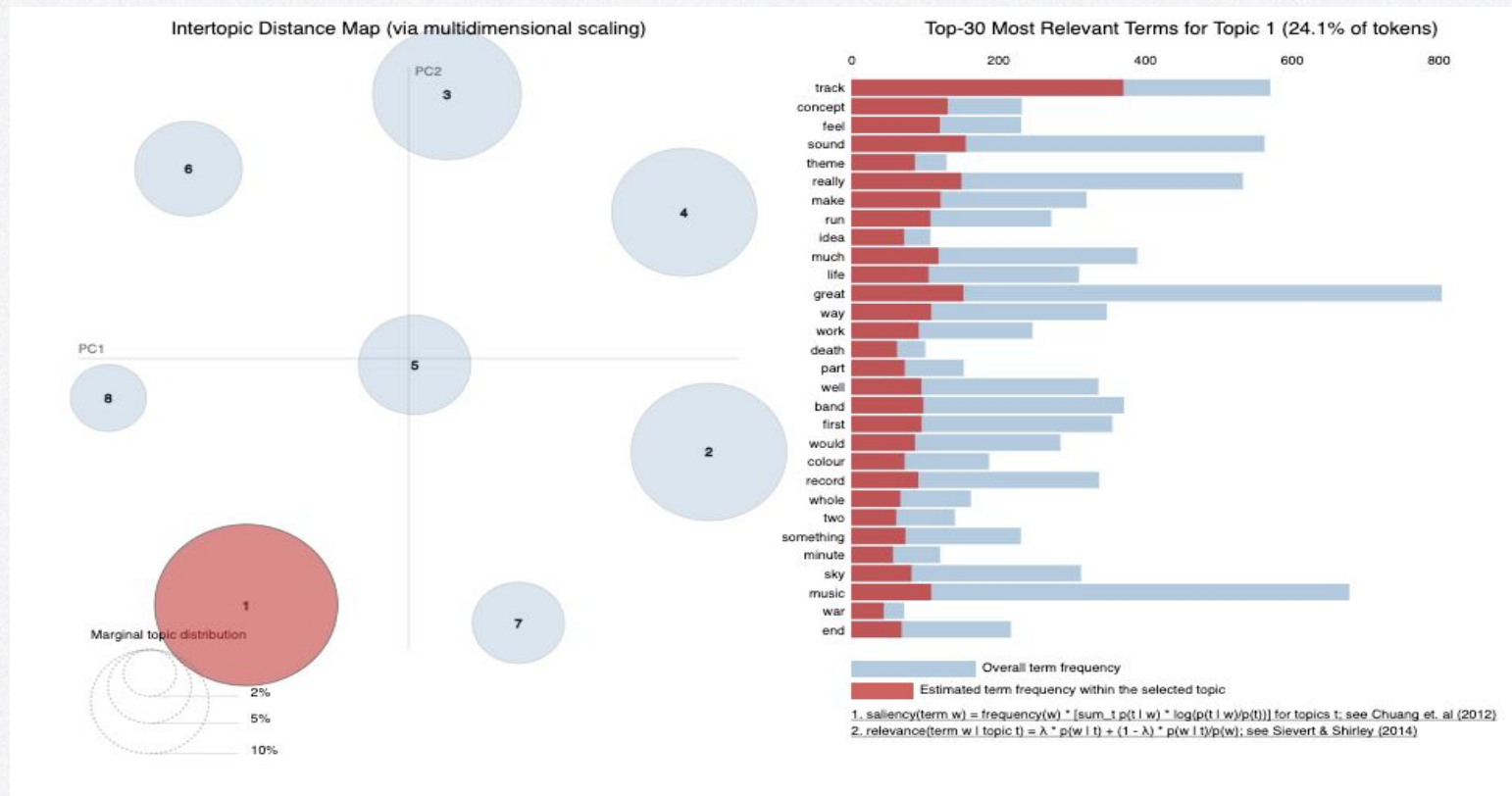
3

Findings – Topic Modeling

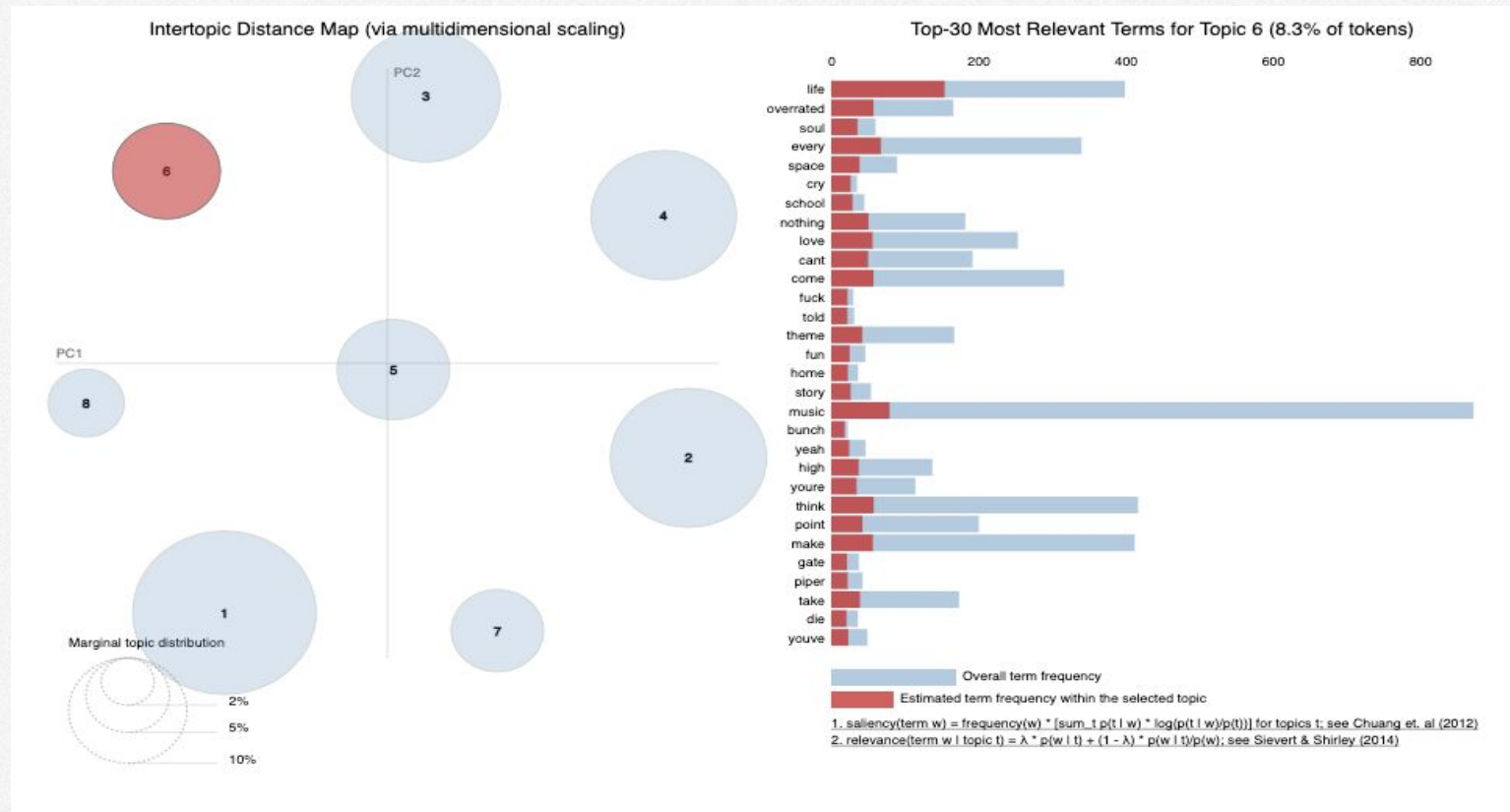
- I determined 8 topics provided the most interpretable results, four of the most salient topics were as follows:
 - a) Extremely **lengthy, highly positive** and **analytical reviews**
 - b) Non-English reviews (Italian, Spanish, and Portuguese reviews were identified)
 - c) Short reviews labeling the album a “**classic**” or the “**greatest album of all time**”
 - d) Reviews noting the album is **too overrated**, or perhaps **good but overrated**



Topic Modeling – LDA Visualization

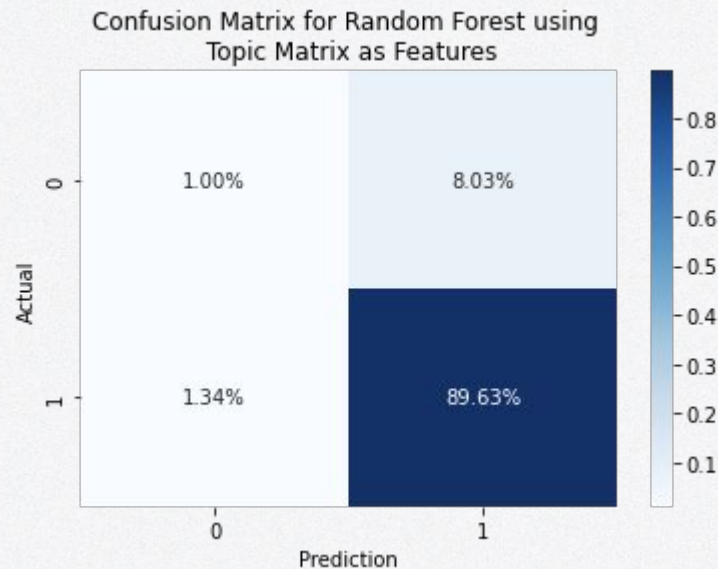


Topic Modeling – LDA Visualization



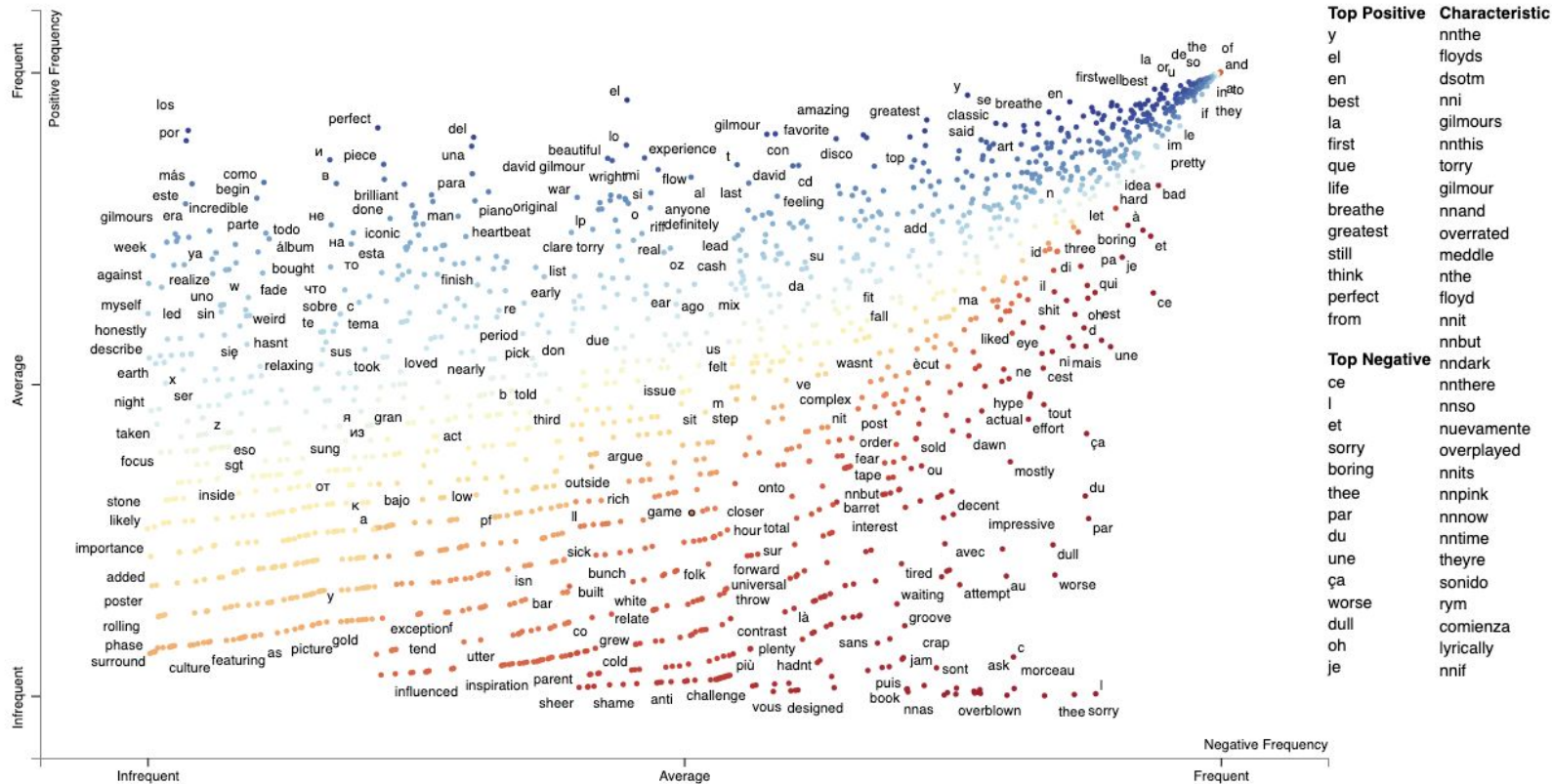
Classification – Random Forest

- A random forest was able to provide the most nuanced results, other models were having **difficulty predicting true negatives**
- This dataset is **extremely unbalanced**, with **largely favorable reviews**
- This model suffers from the same issue, however, it shows promise as far as its ability to **capture** some of the **true negatives**



- Accuracy: **0.904**
- Recall: **0.975**
- Precision: **0.923**
- F1: **0.948**

Scatter Text – Sentiment Analysis



Positive document count: 1,376; word count: 189,462

Negative document count: 118; word count: 17,834

Conclusions

- 1) Dark side of the moon is widely regarded as one of the **“greatest”** albums by reviewers
- 2) There is speculation that the album is **“overrated”**
- 3) The **“guitar”**, **“bass”**, and **“piano”** were mentioned frequently in reviews, seemingly well-regarded
- 4) The reviews were rather **homogeneous**

Recommendations & Next Steps

- 1) Gather **more data** – this will allow clearer topics to emerge
 - a) Analyzing **full-bodies of work** of **multiple artists within a single genre**
- 2) Utilize **CorEx** to hone in on specific topics
- 3) Use of genre data
 - a) Perhaps construction of a content-based **recommendation system**

Thank you!

