A Musician's Alter Ego

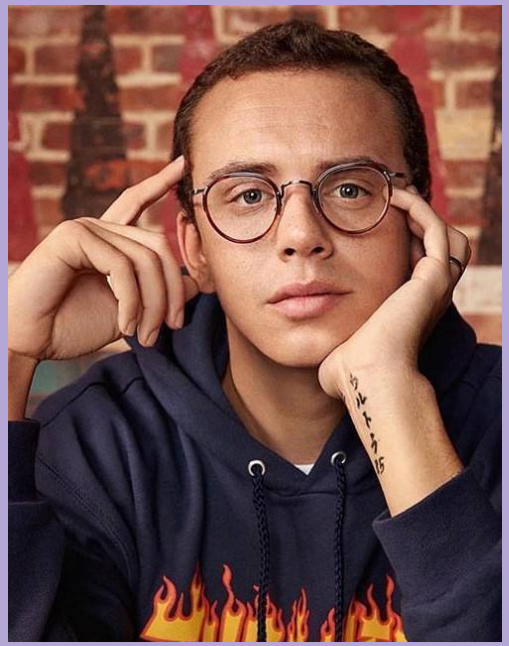
Benjamin Lin

June 1st, 2018


camille

Robert Bryson Hall II

Logic

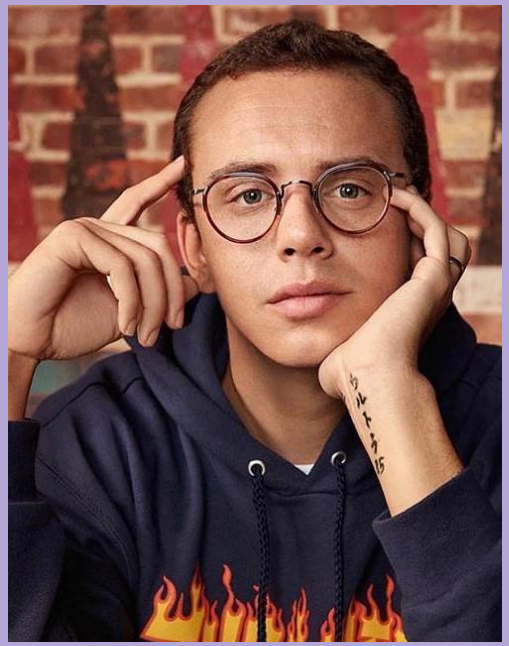


Young Sinatra

Bobby Tarantino

Robert Bryson Hall II

Logic



Peace, love, society

Young Sinatra

Success, motivated, vengeful

Bobby Tarantino

Outspoken, turn-up, carefree

Objectives

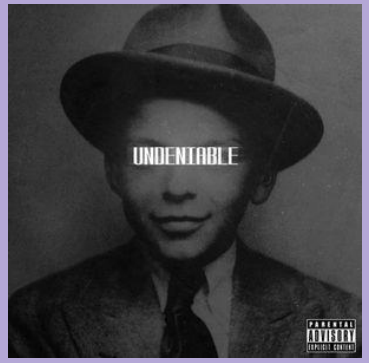
1. Cluster his lyrics into different groups
2. Predict if a line came from Logic, Young Sinatra, or Bobby Tarantino

Data

Logic

2,100 rows

Young Sinatra



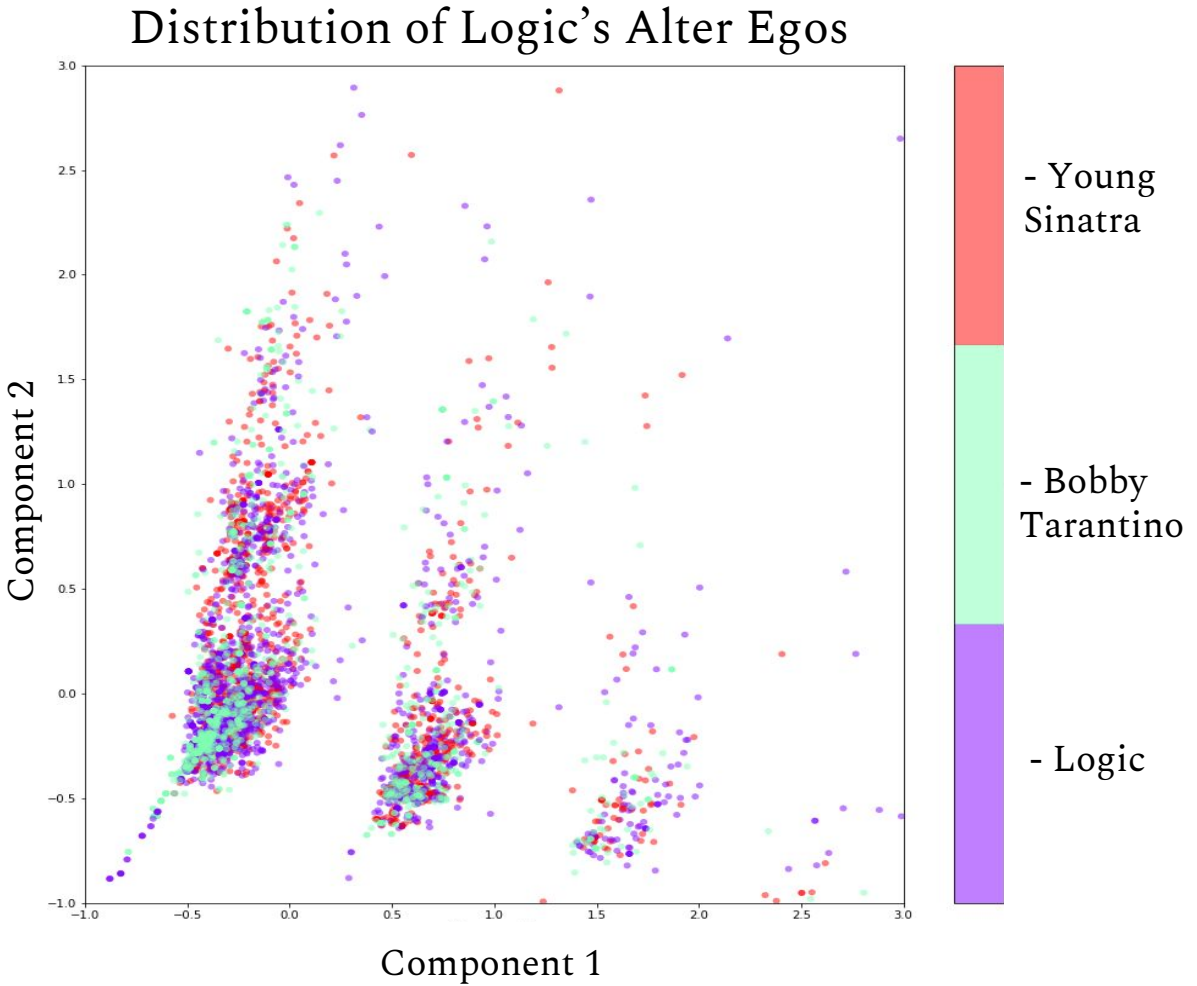
2,000 rows

Bobby Tarantino

1,800 rows

Unsupervised Clustering

PCA


Distribution of Logic's Alter Egos
Component 2
Component 1
3.0
2.5
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.0
−0.5
0.0
0.5
1.0
1.5
2.0
2.5
3.0
- Young Sinatra
- Bobby Tarantino
- Logic

LDA

Alter Ego Topics

Let's revisit his 3 alter-egos and their corresponding topics:

Logic	Young Sinatra	Bobby T.
- Peace - Love - Society	- Success - Motivated - Vengeance	- Outspoken - Turn-up - Carefree

LDA Topics

LDA best split the text into 2 topics. Here are the top terms for each topic:

Topic 1	Topic 2
- Everybody - Love - Mind - People - World	- Money - Nobody - Livin - B*tch - Give

LDA Topics

LDA best split the text into 2 topics. Here are the top terms for each topic:

Topic 1	Topic 2
- Everybody - Love - Mind - People - World	- Money - Nobody - Livin - B*tch - Give
Logic!	Young Sinatra/Bobby T!

Supervised Classification

Models

Features

Algorithms

	Bag of Words	tf-idf	Spacy Features	Word2Vec
Multi N.B.	0.672	0.674	0.427	0.417
KN Classifier	0.604	0.646	0.532	0.592
SVC	0.601	0.429	0.429	0.574

*metric = test set accuracy score


n [48]: who_said(['i want the world to be a better place'])
['Logic']
n [49]: who_said(['lets drink and party all night'])
['Bobby Tarantino']
In [ ]: who_said([''])

The End!

Email: benlin100@gmail.com
LinkedIn: linkedin.com/in/benjaminlin24/
Github: github.com/benlin100

Appendix

Agenda

1. Background & Objectives
2. Data
3. Unsupervised Clustering
4. Supervised Classification
5. Conclusion

Background & Objectives

Summary

1. Cluster his lyrics into different groups
    - **PCA**
    - **LDA (Topic Modeling)**
2. Predict if a line came from Logic, Young Sinatra, or Bobby Tarantino
    - Features: Bag of Words, SPACY, W2V, tf-idf
    - Modeling Algorithims: Multi N.B, KN Classifier, SVC

Models

Features

Algorithms	Bag of Words	tf-idf	Spacy Features	Spacy + BoW	Word2Vec
Multi N.B.	0.672	0.674	0.427	0.665	0.417
KN Classifier	0.604	0.646	0.532	0.577	0.592
SVC	0.601	0.429	0.429	0.429	0.574

*Metric = Test set accuracy score