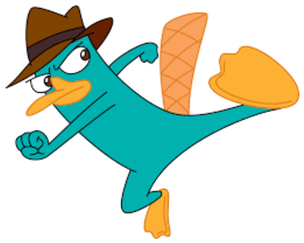


Ghost Writer Detector

Statistical Significance Score of Lyric Similarity

Nicholas Kim

UNIVERSITY OF MICHIGAN

Introduction

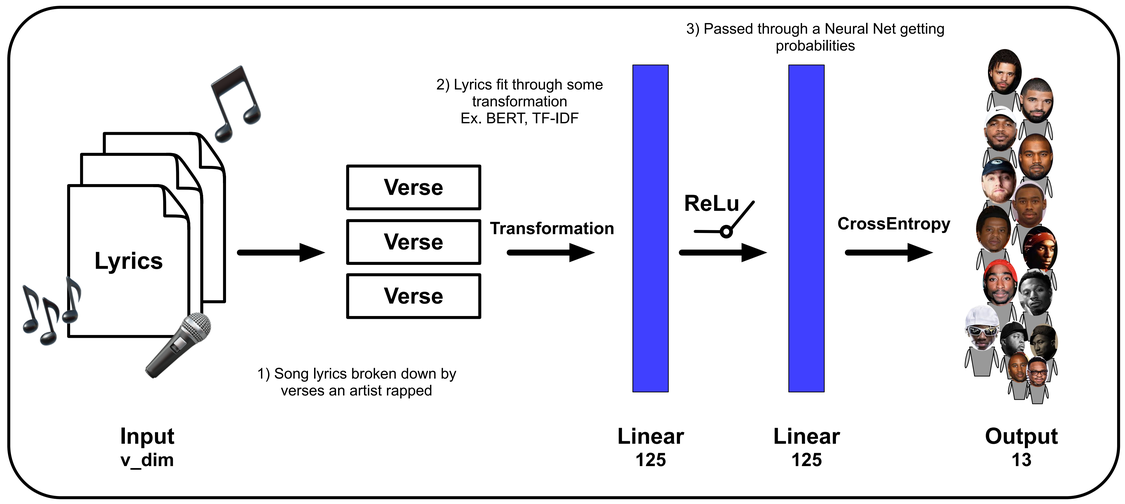
Rapper's and MC's take pride in their writing chops making any allegation of ghostwriting to be head line news. The most public of which being between Drake and Quentin Miller. While these allegations are only made possible through insider leaks, is there a way we can detect these instances of bar malpractice quantitatively? This project will offer such a solution by creating a score that judges how statistically similar a songs lyric is between two artists using classification probabilities and bootstrapping methods

Methodology

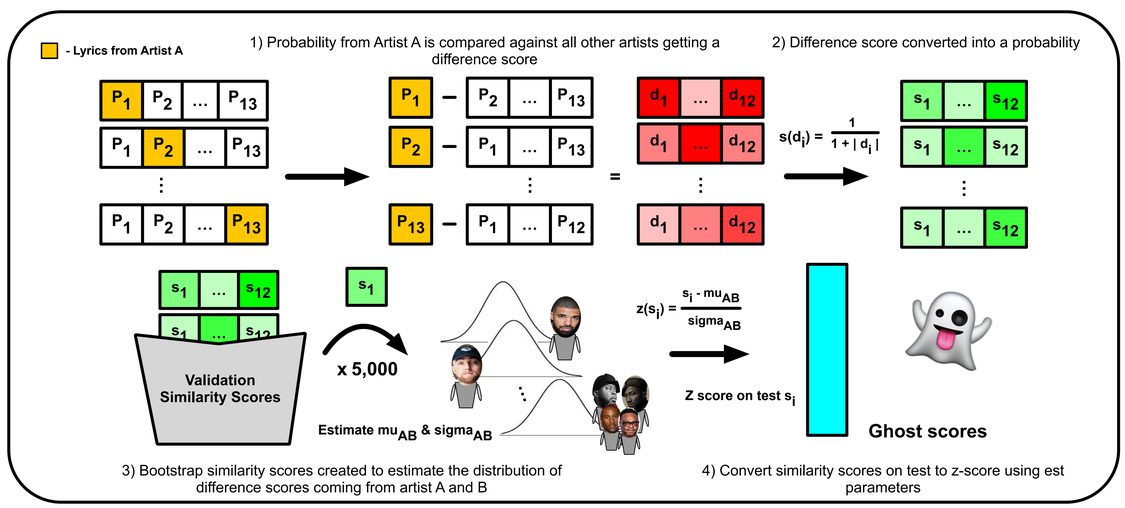
Model Classifier:
Lyrics were first split into just verses that they rapped and passed into TF-IDF transformation to get a higher dimensional representation. These vectors were then passed into a Deep Neural Network to get probabilities of a song coming from 1 of the 13 artists in our corpus.

Ghost Writer Score:
Now that we have a measure of how likely a songs lyric is from an artist we can use it to find those who had a similar score. To judge how statistically significant that difference measure if we can take a standardized score where the means and standard deviations were found using a bootstrapping method.

Model Classification

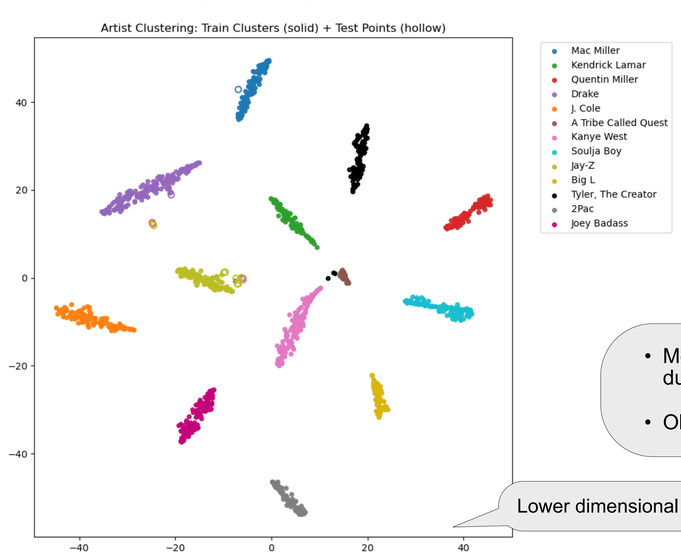


Ghost Score



Results

Classification Visual



Lower dimensional representation of how songs were grouped on both train and test

Classification Accuracy

Model	Val Accuracy	Test Accuracy	F1 Score
Random Guess	20%	7%	15.16%
BERT	27.33%	20%	15.28%
TF-IDF	59.33%	46.67%	30.83%
TF-IDF SGD	67.66%	86.67%	74.29%
TF-IDF SGD Big	30.67%	26.67%	12.04%

- Model had difficulty classifying Kanye West in both test and val. Likely due to his high writing collaboration with other artists
- Old school rappers like 2pac and Big L were easier to classify

- 6 Man and 10 Bands were public cases of ghost writing allegations! Our model similary flagged these and correctly identified Quentin Miller!

Ghost Scores on Test

Song Title	Lyrics By	Ghost Score	Likely Ghost Writer
Izzo	Jay-Z	1.99	Kanye West
6 Man	Drake	4.03	Quentin Miler
10 Bands	Drake	3.25	Kanye West
Facts	Kanye West	0.33	Jay-Z
Legends	Drake	0.78	Jay-Z

Future Work

- More false positive analysis. Are instances where artists are inspired by another lyricist being unfairly flagged in our model? Look into how sensitive our score is
- Era analysis. Are there more instances of flagged ghost writing now or before?