



THE DATA & MUSIC COMPANY



HOW TO MAKE YOUR SONG A "BOP"

by Christian Moya, Data Scientist

PRESENTATION OUTLINE

OUR DISCUSSION POINTS

About Us

Problem Statement

Methodology

Data

Results

Recommendations

Future Work

OUR VISION

We are a group of data scientists that help independent artists excel in the music industry using data-driven recommendations.



DO LYRICS HAVE AN EFFECT ON ITS SUCCESS IN THE BILLBOARDS?

NULL: LYRICS **DO NOT** HAVE AN
EFFECT ON ITS SUCCESS.

ALTERNATE: LYRICS **DO** HAVE AN
EFFECT ON ITS SUCCESS.



Methodology - CRISP-DM



BUSINESS UNDERSTANDING

Lyricality, how are the top 25 songs different than the rest?



DATA UNDERSTANDING

Billboard Top 100 (1970-2020) and Lyrics from Genius (2,000+ songs)



DATA PREPARATION

Get the data ready

MODELING

Random Forest Classifier
Naive Bayes Classifier

EVALUATION

How can we make our models better?

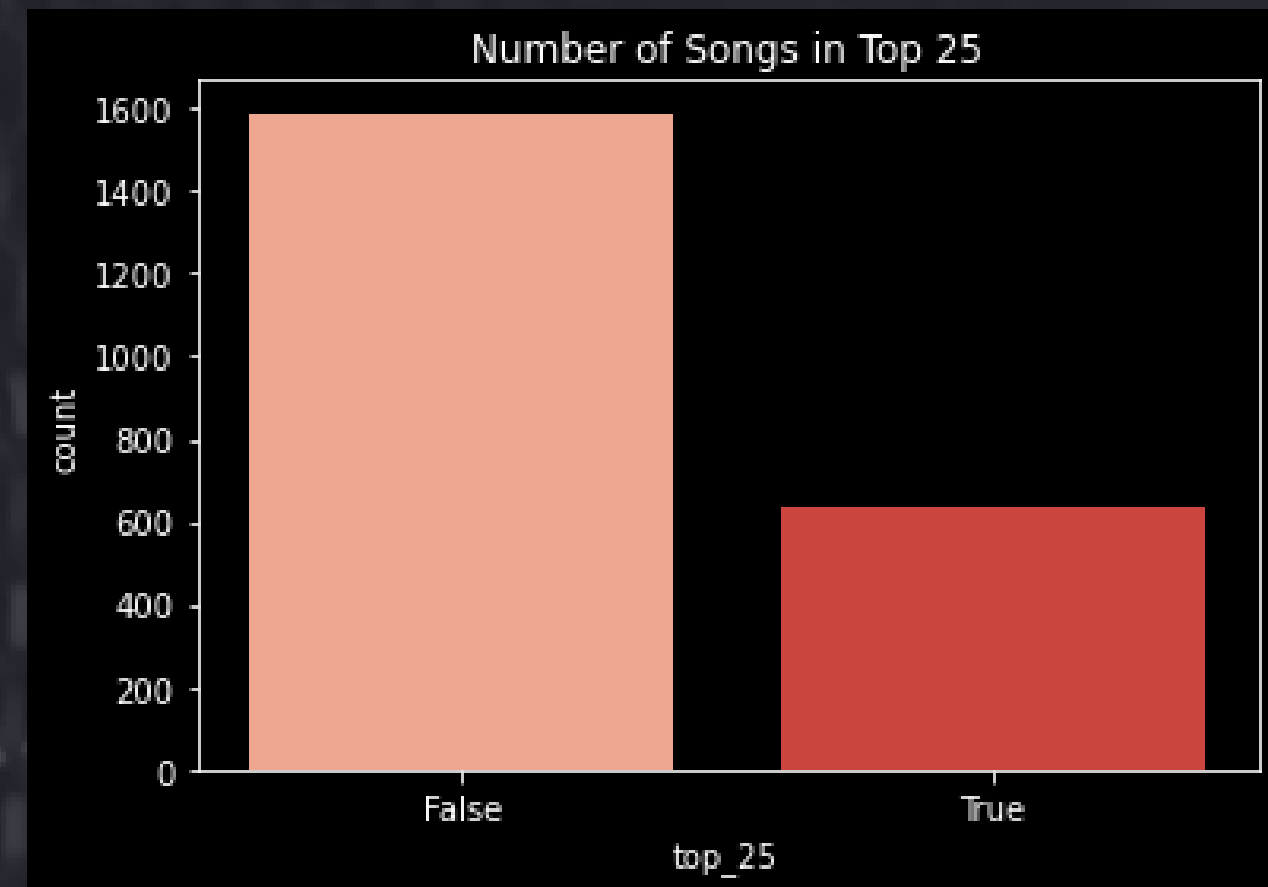
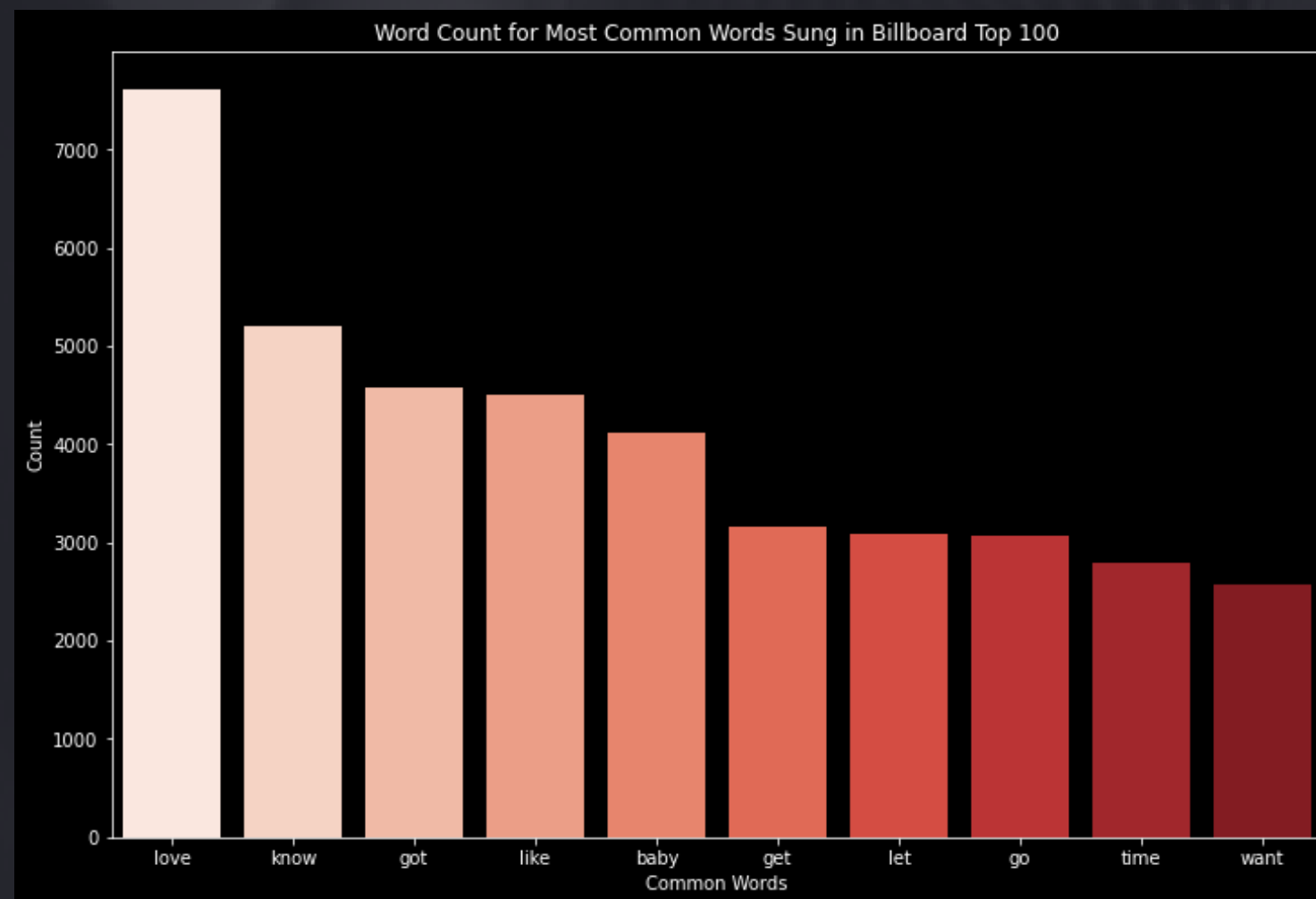
DEPLOYMENT

Recommendations to Follow

EXPLORATORY DATA ANALYSIS

THE DATA

Over 2,200 songs from 1970 - 2020
Most commonly used word: "Love"



TOP ARTISTS

19: Madonna
18: Elton John, Taylor Swift
16: Drake
15: Rihanna

RESULTS

LITTLE IMPACT

Baseline:

67% for Naive Bayes

67% for Random Forest

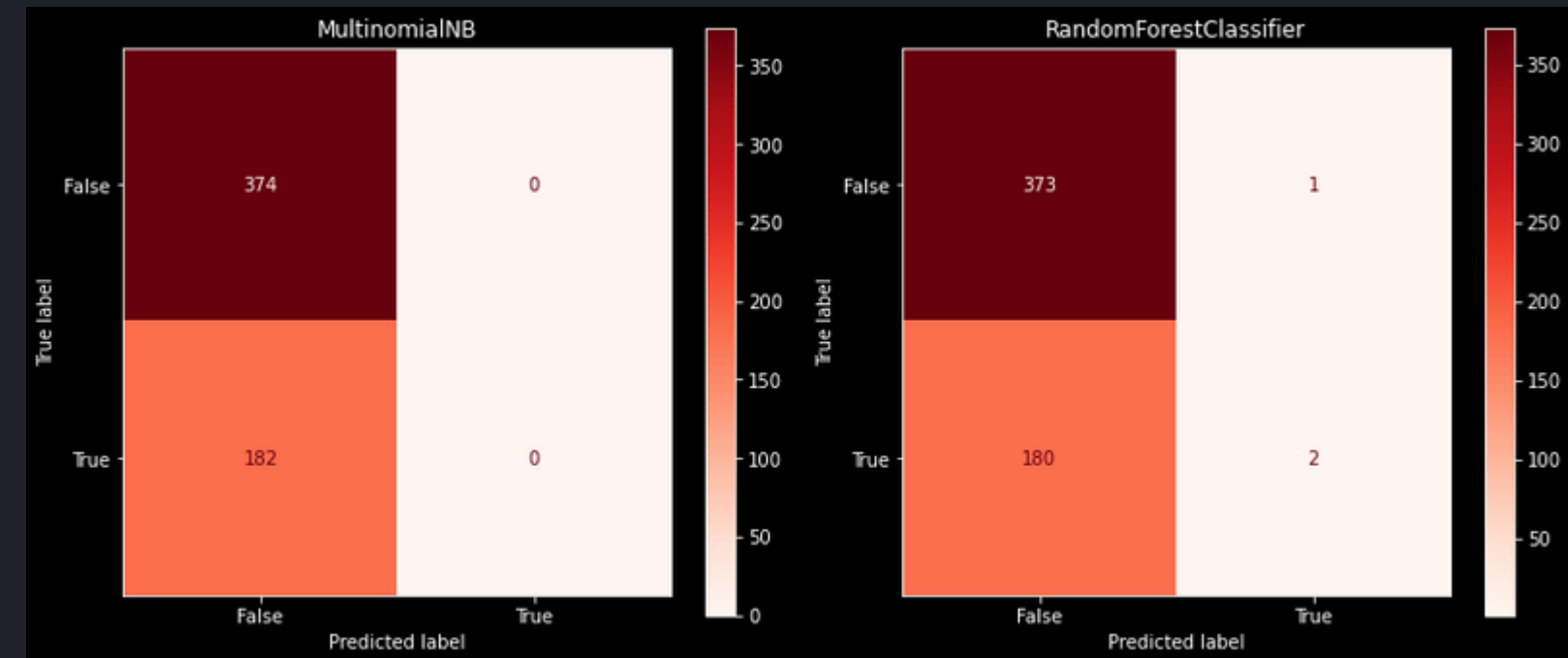
Best Model Performance:

76% for Naive Bayes

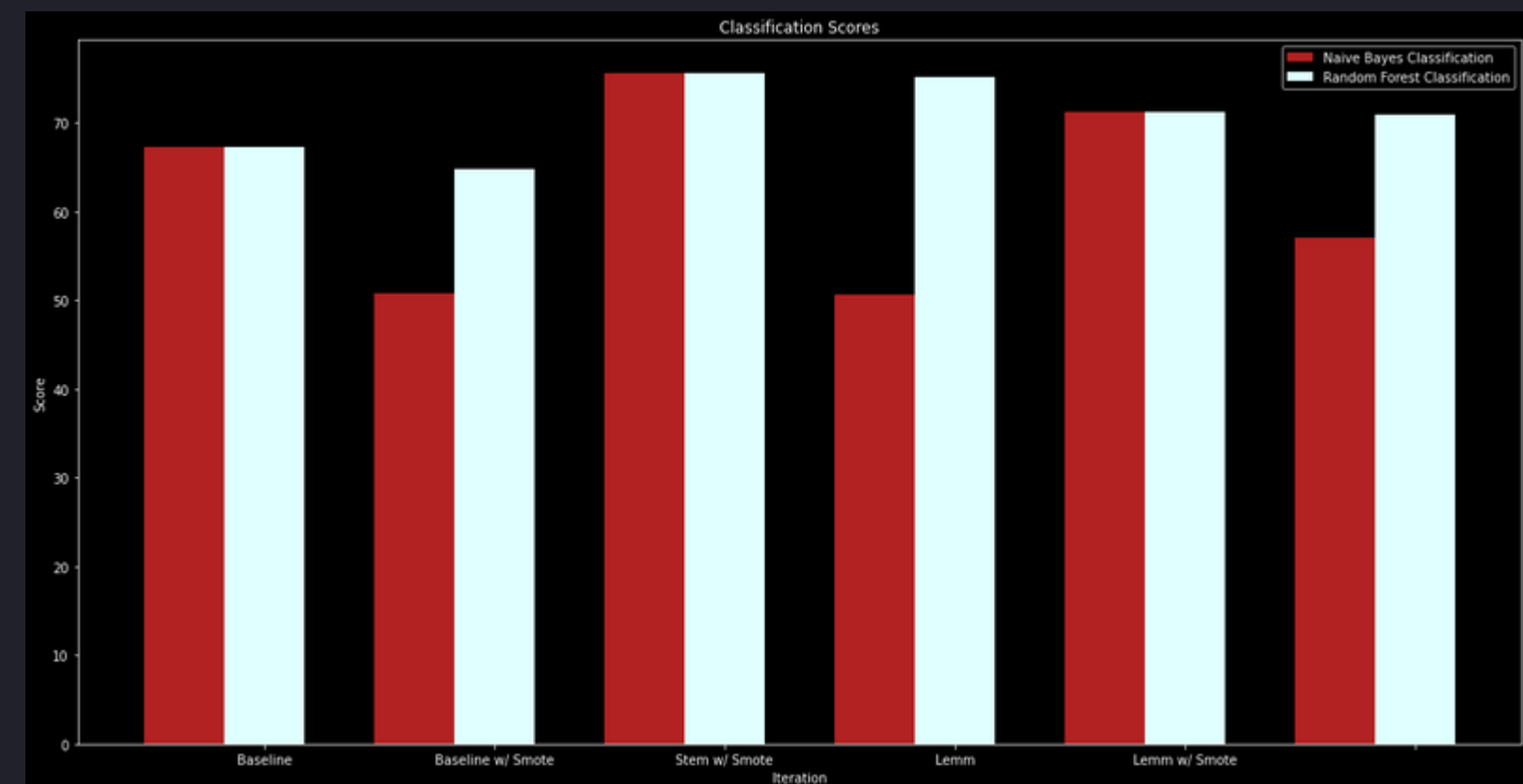
76% for Random Forest

- Results are just as good as saying False for every prediction, therefore we fail to reject our null hypothesis.

CONFUSION MATRIX FOR BASELINE MODEL



ACCURACY SCORES FOR CLASSIFICATION MODELS

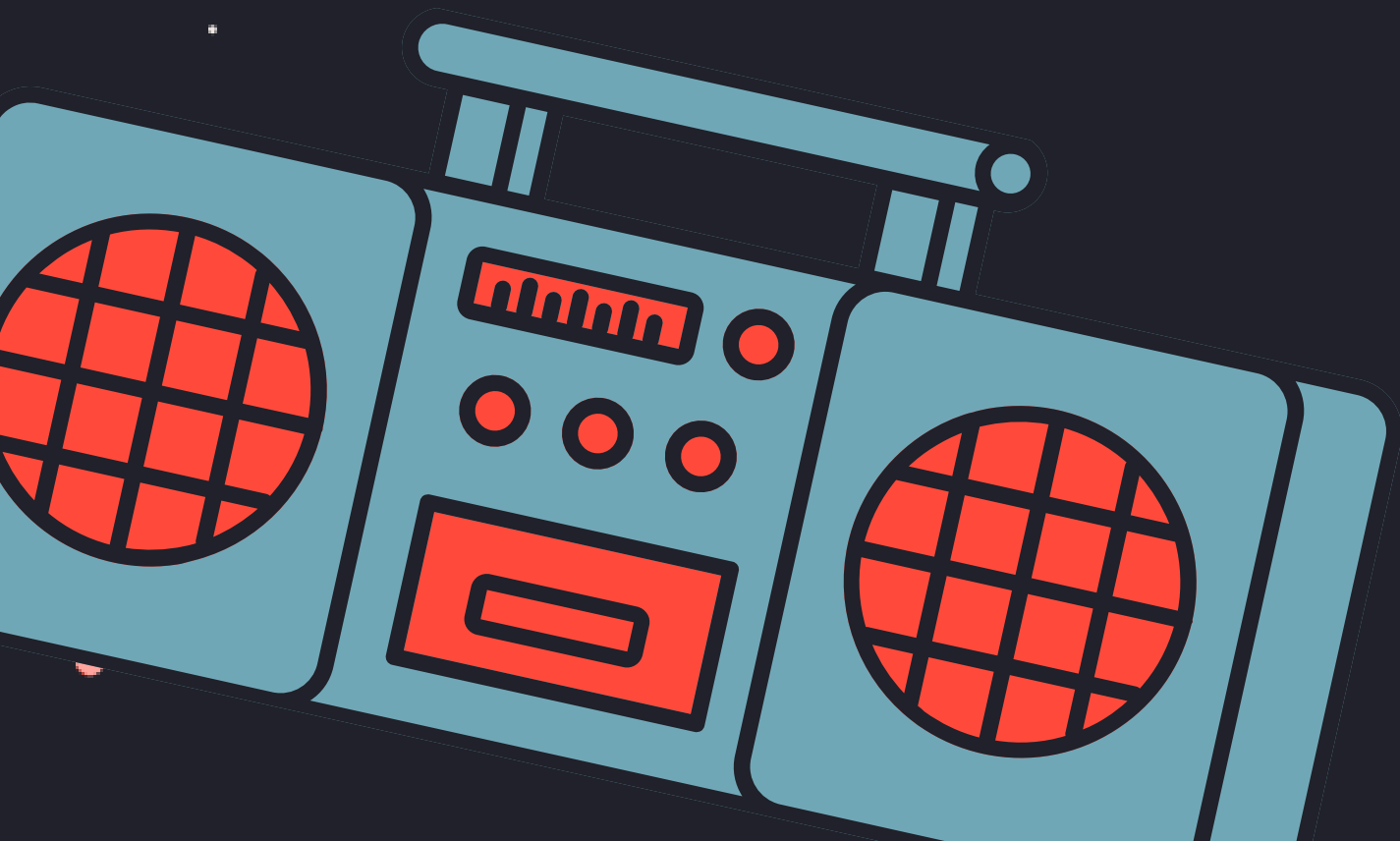


RECOMMENDATIONS

1

FOCUS ON THE BEAT OF THE SONG

Words have little impact on whether a song will make the top 25 or not, so the tempo and musical composition of the instruments may have more influence on whether a song is a hit or not.

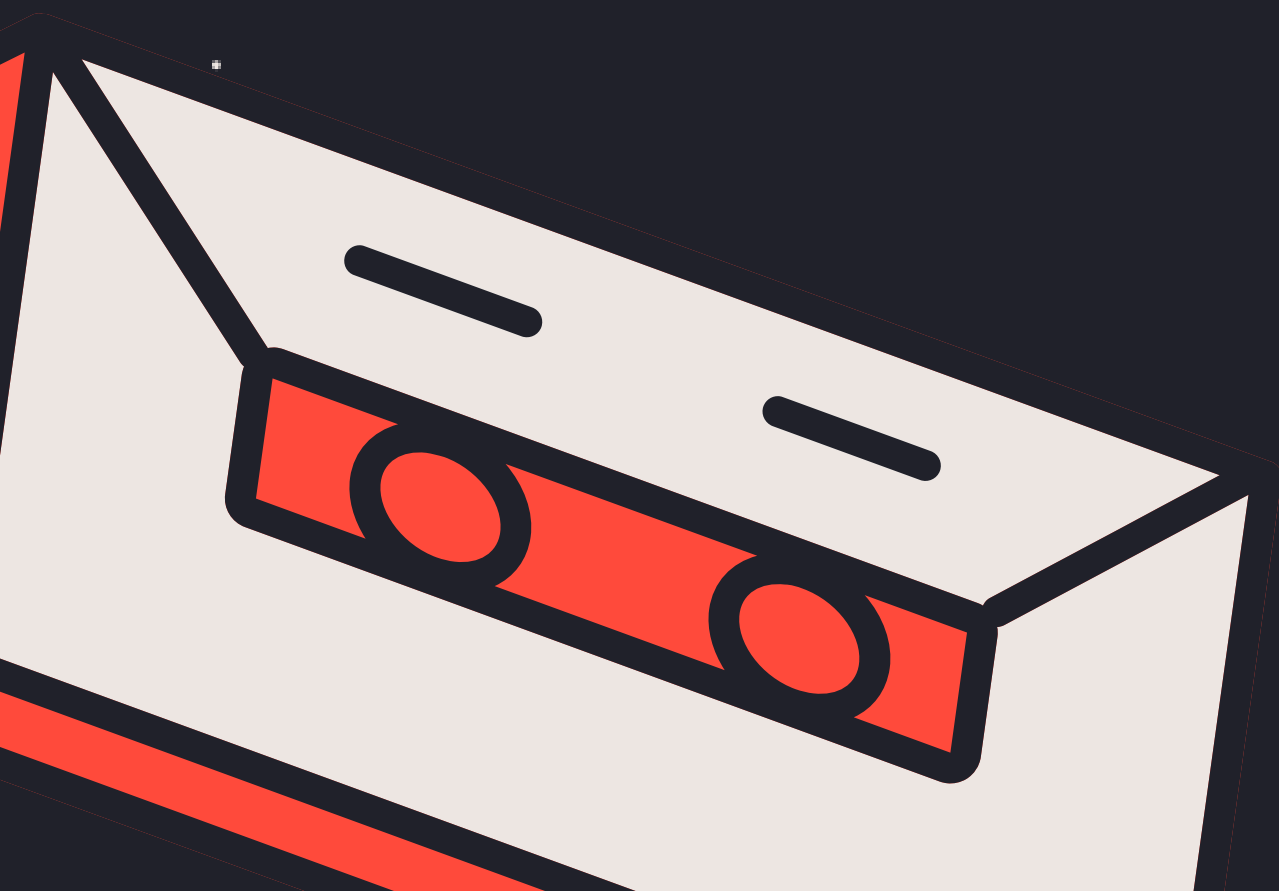


RECOMMENDATIONS

2

WRITE ABOUT RELATABLE TOPICS

Though our model could not predict if a song is in the Top 25 or not, we do know that each song in our database is a Billboard Top 100. That data also shows the most common word is "love", a concept many have strong feelings for. The second most common word is "know".



RECOMMENDATIONS

3

MAKE A SONG ABOUT ANYTHING YOU WANT TO SING ABOUT

Because our model could not predict a song's success based on words alone, it shows that songs about dancing would do just as fine as a song about a breakup. Write what you want to write about.





Future Work

SCRAPE MORE SONG LYRICS

My data only contained songs from Billboard Top 100. Song lyrics that aren't hits could help classify better.

ADD ADDITIONAL FEATURES

Songs are so much more than its lyrics, it's a combination of sound and tempo that make it the creation it is. Additional features can help build a better model.

CLASSIFY FOR GENRES

Do certain words appear more in certain genres? We'll find out next time.

FINAL WORDS



Music is the
movement of
sound to reach
the soul for the
education of its
virtue.

PLATO

Thank you!

CONTACT US

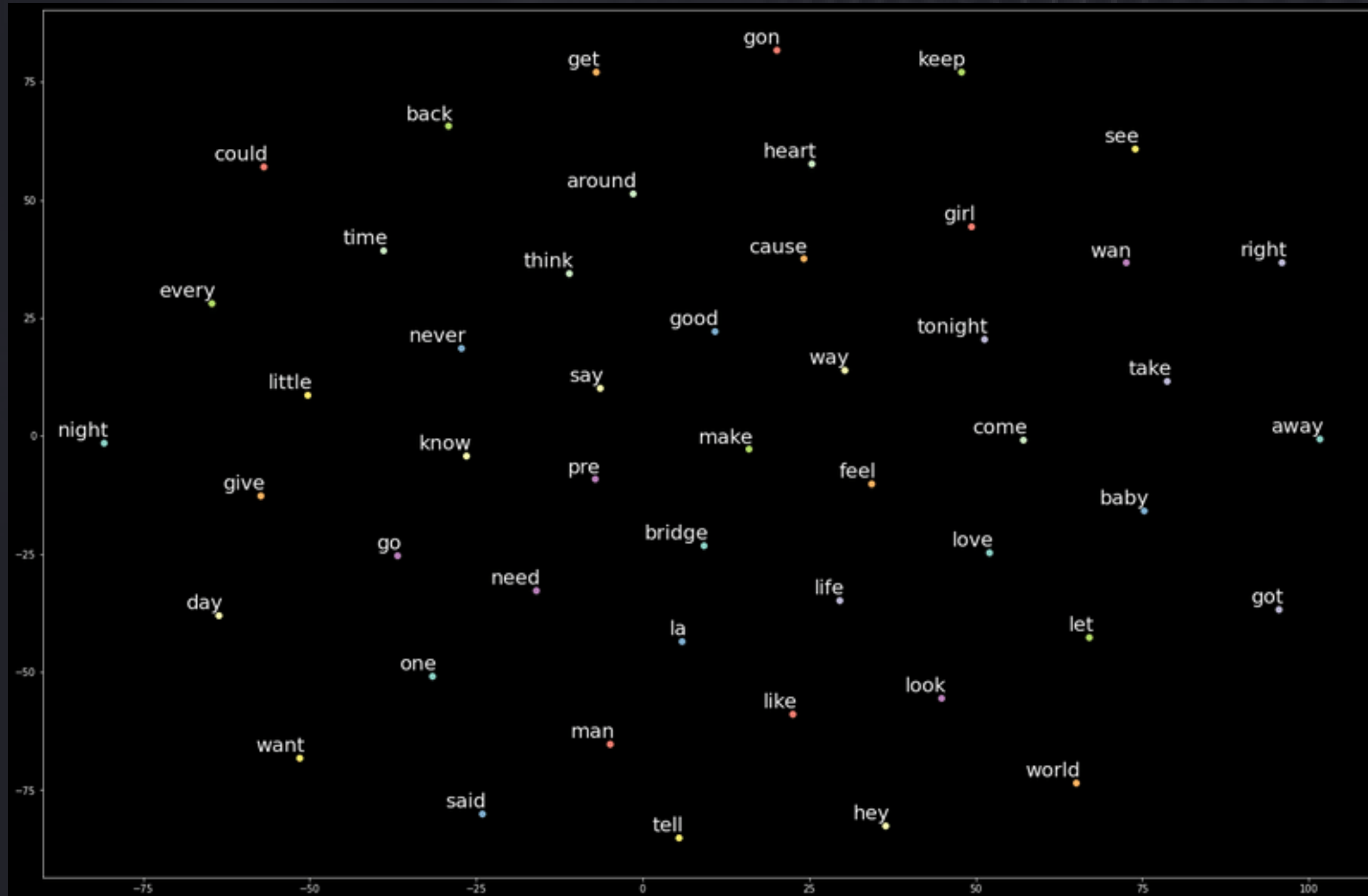
WE'D LOVE TO HEAR FROM YOU!

EMAIL ADDRESS

christian.r.moya@gmail.com

APPENDIX

T-DISTRIBUTED STOCHASTIC NEIGHBOR EMBEDDING



LATENT DIRICHLET ALLOCATION

Topics found by bag of words LDA

Topic #1: baby love take whoa

Topic #2: know back gon go

Topic #3: night way eyes like

Topic #4: got want like bad

Topic #5: la good like got

Topic #6: get dance tonight let

Topic #7: like got hey go

Topic #8: love know time got

Topic #9: like little ya let

Topic #10: let doo baby go

Topics found by tf-idf LDA

Topic #1: gon come baby take

Topic #2: love heart baby know

Topic #3: keep got eyes like

Topic #4: need want say one

Topic #5: like night think feel

Topic #6: away day take say

Topic #7: go back life time

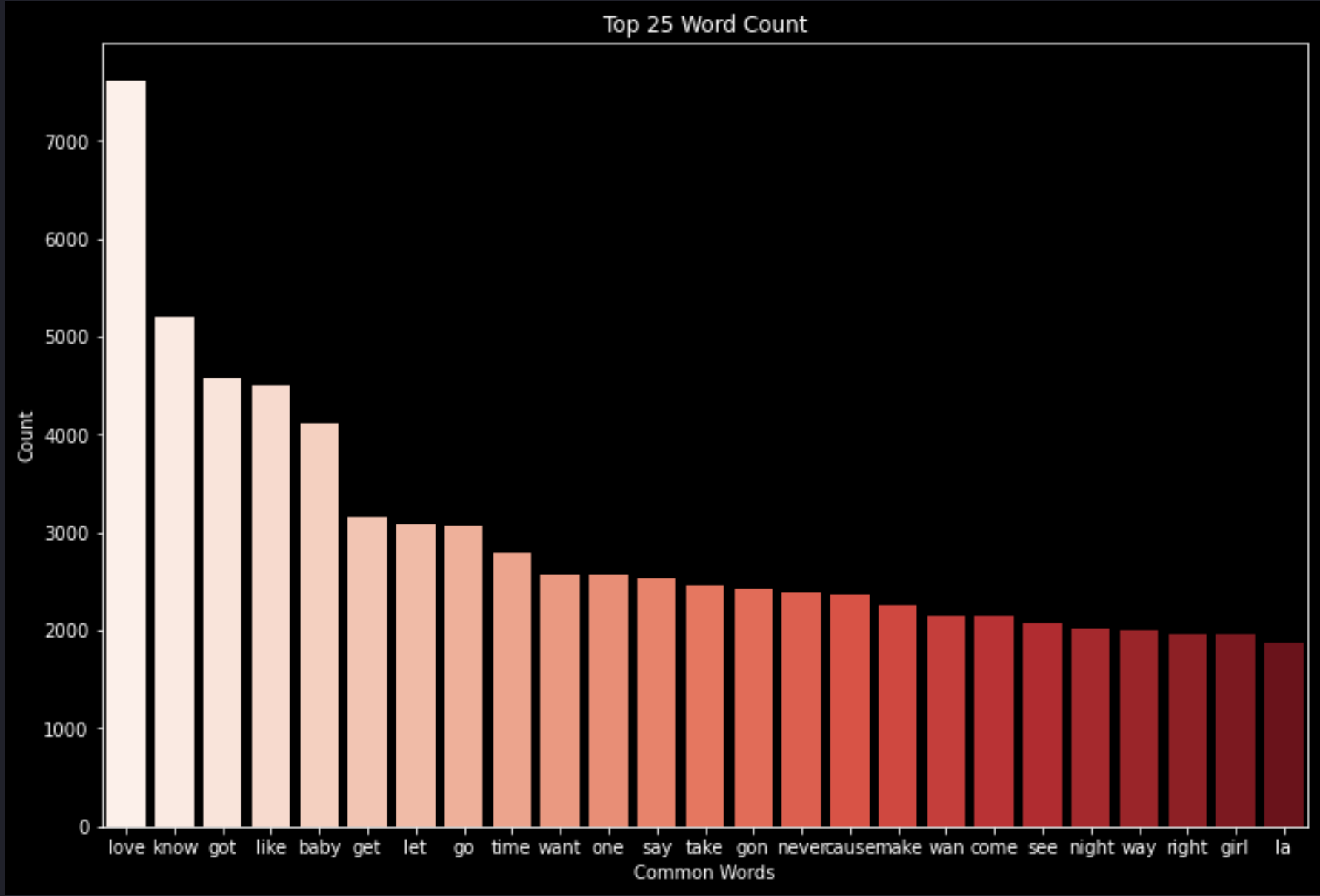
Topic #8: wan get got give

Topic #9: baby girl good love

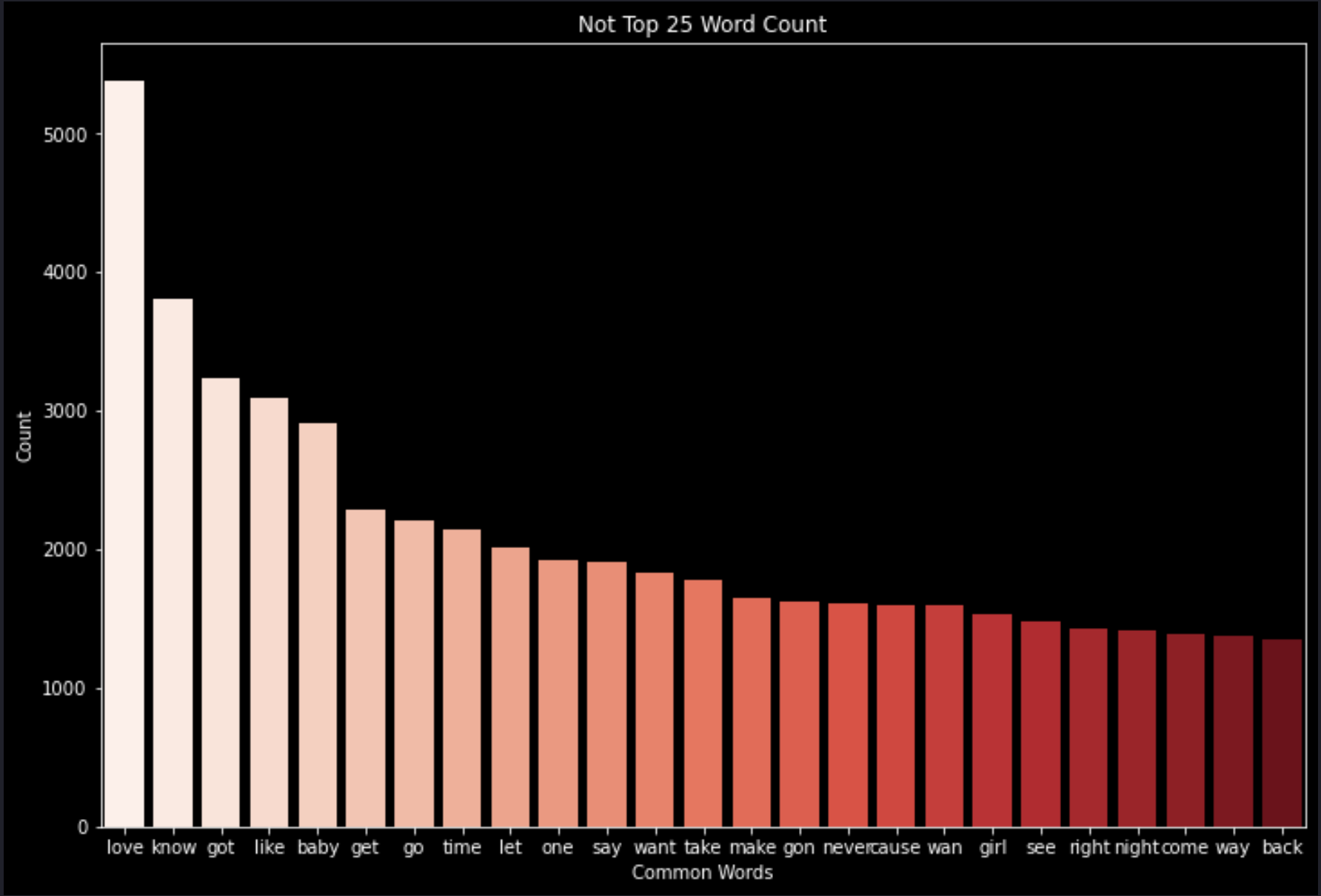
Topic #10: never could let know

APPENDIX

TOP 25

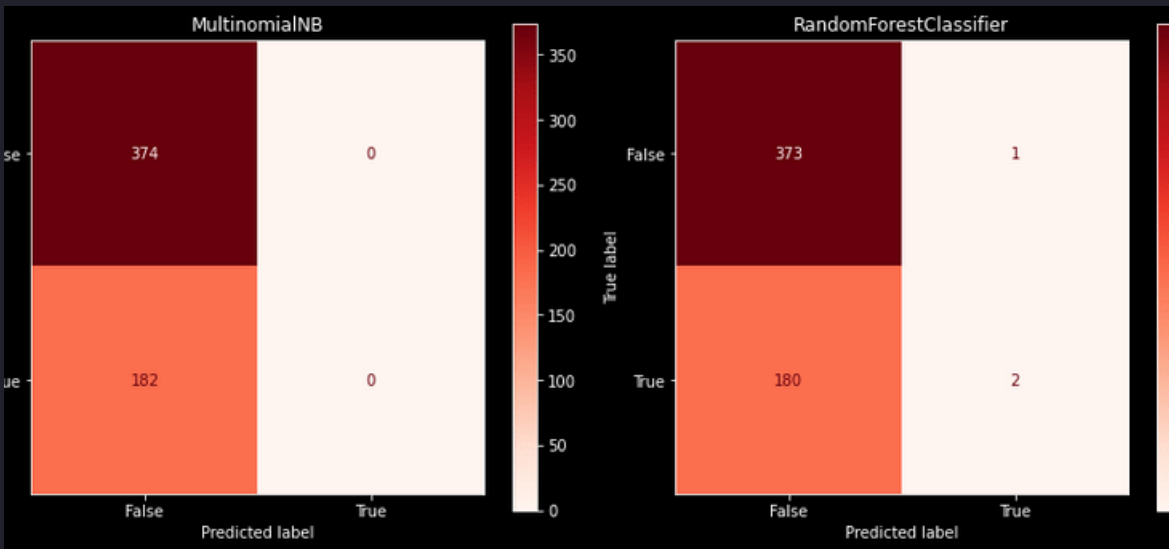


NON TOP 25

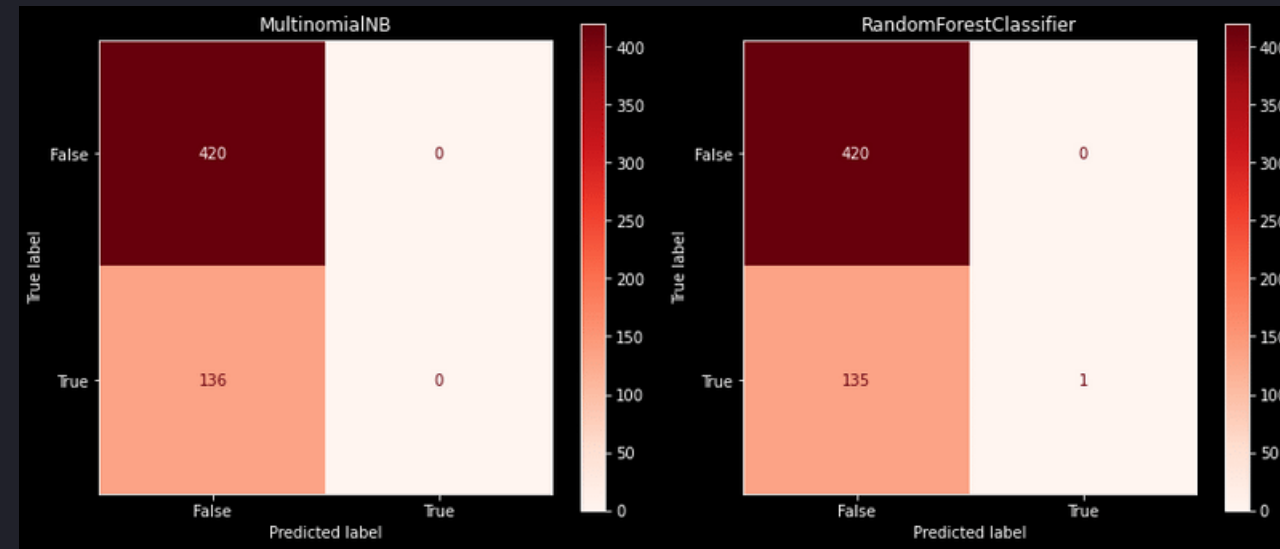


APPENDIX

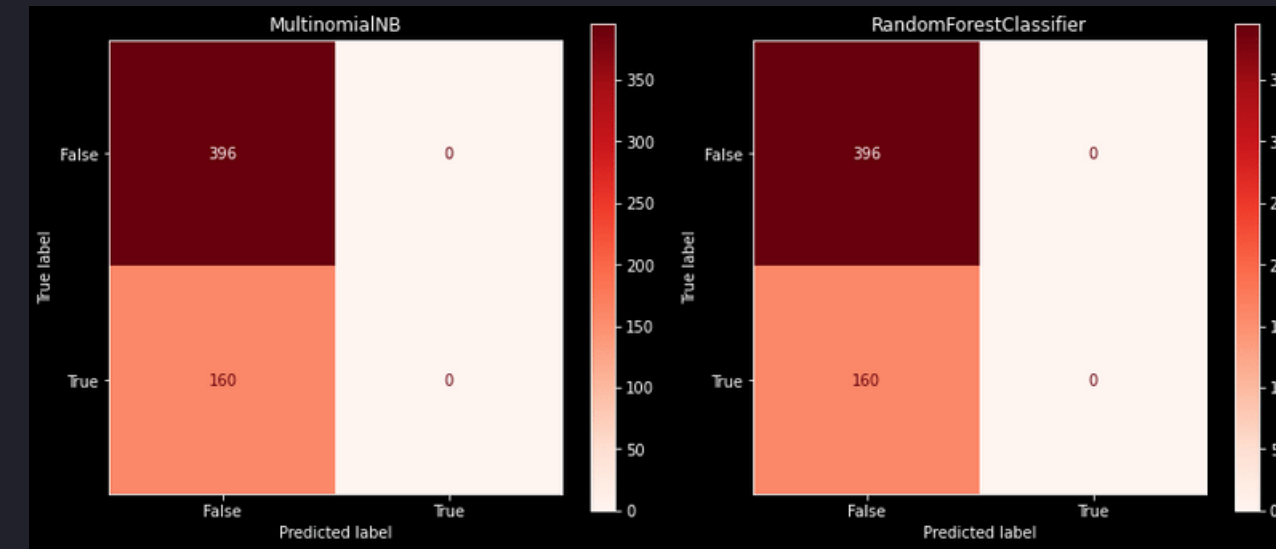
BASELINE MODEL



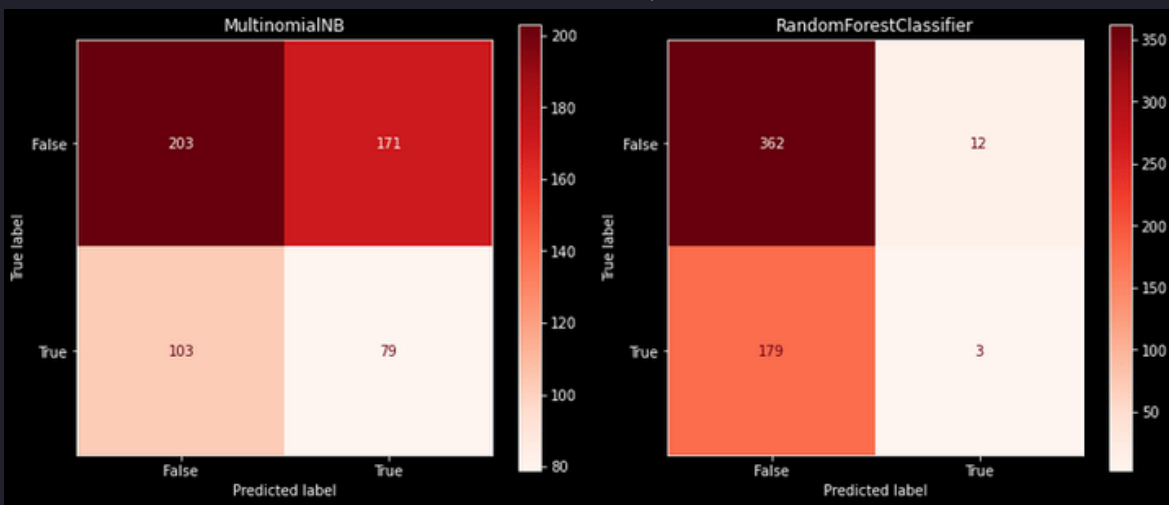
STEMMATIZED MODEL



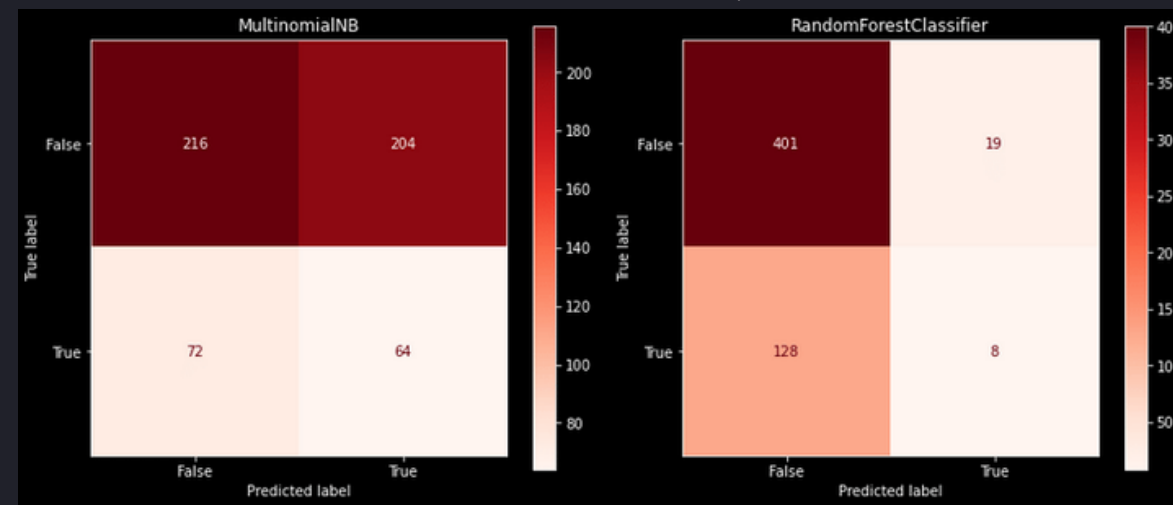
LEMMATIZED MODEL



BASELINE W/ SMOTE



STEMMATIZED W/ SMOTE



LEMMATIZED W/ SMOTE

