PREDICTING MY MUSIC "TYPE"

USING PERSONAL SPOTIFY LISTENING DATA TO PREDICT THE SONGS THAT MAKE MY FAVORITES PLAYLIST

Austin Heath

STAT 426
Data Science Methods
Final Project

December 10, 2020

Introduction

Part 1

Austin Heath

STAT 426
Data Science Methods
Final Project

December 10, 2020

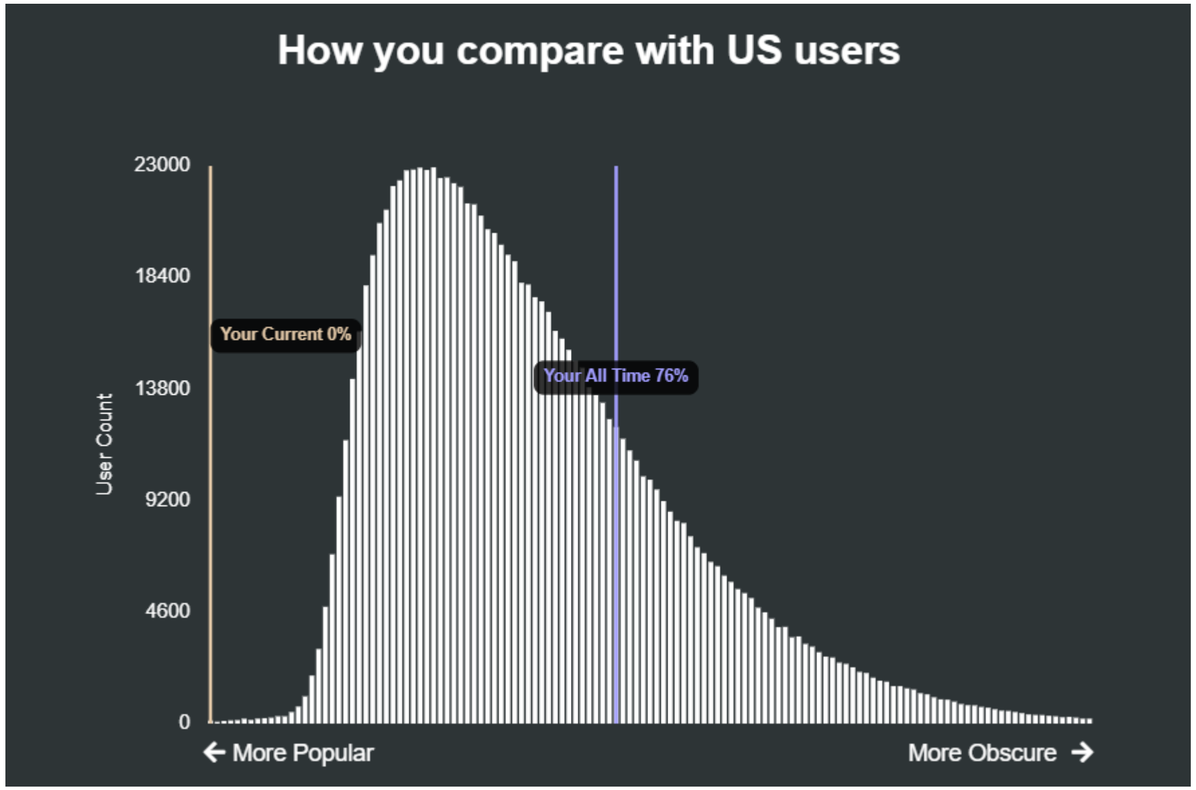
Background

What's My Music Type?

76%

Your music is more obscure than
76% of **1,042,515 US** users on Obscurify.

2,373,613 Global Users



Austin Heath

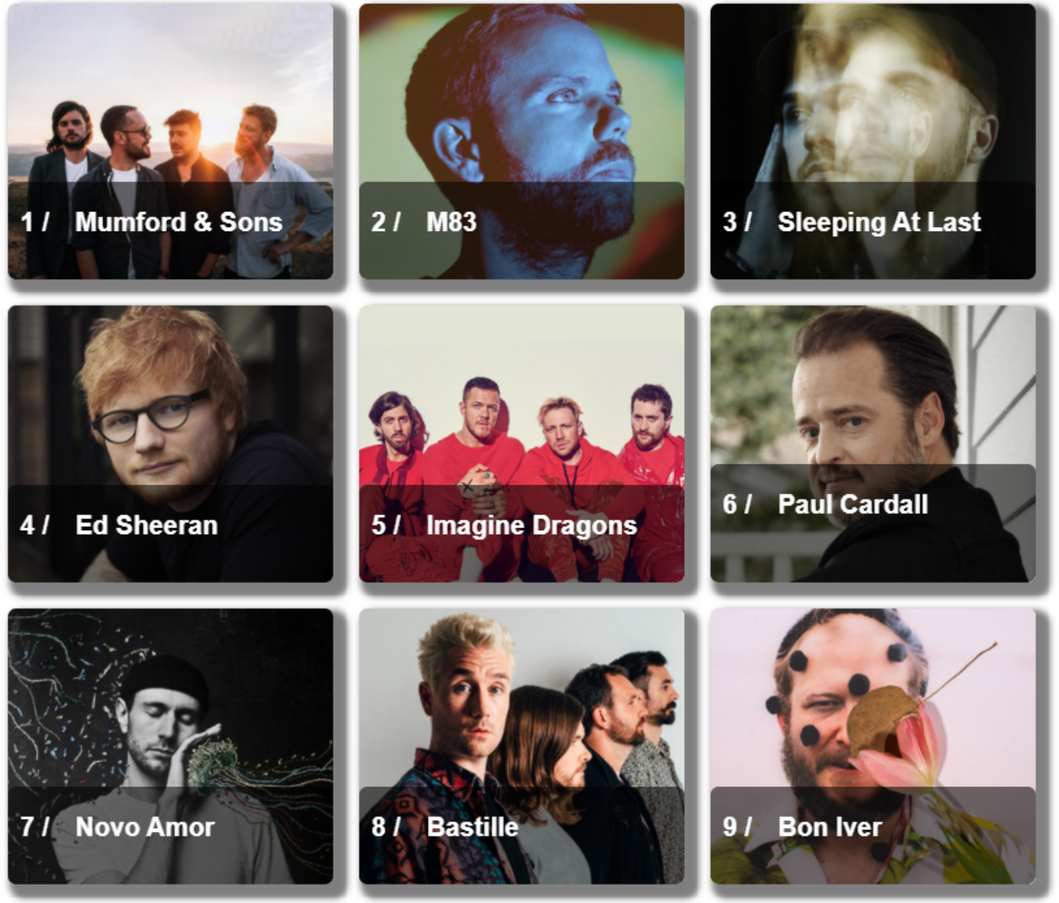
STAT 426
Data Science Methods
Final Project

December 10, 2020

Spotify 2020 "Wrapped"

My Top Songs of the Year

All Time Top Artists



Austin Heath

STAT 426
Data Science Methods
Final Project

December 10, 2020

Objectives

- Using training data from my personal Spotify listening history in 2020, can I accurately classify whether or not songs are in my top 100 songs of the year?
- Using this model, can I recommend new favorite songs?

Process

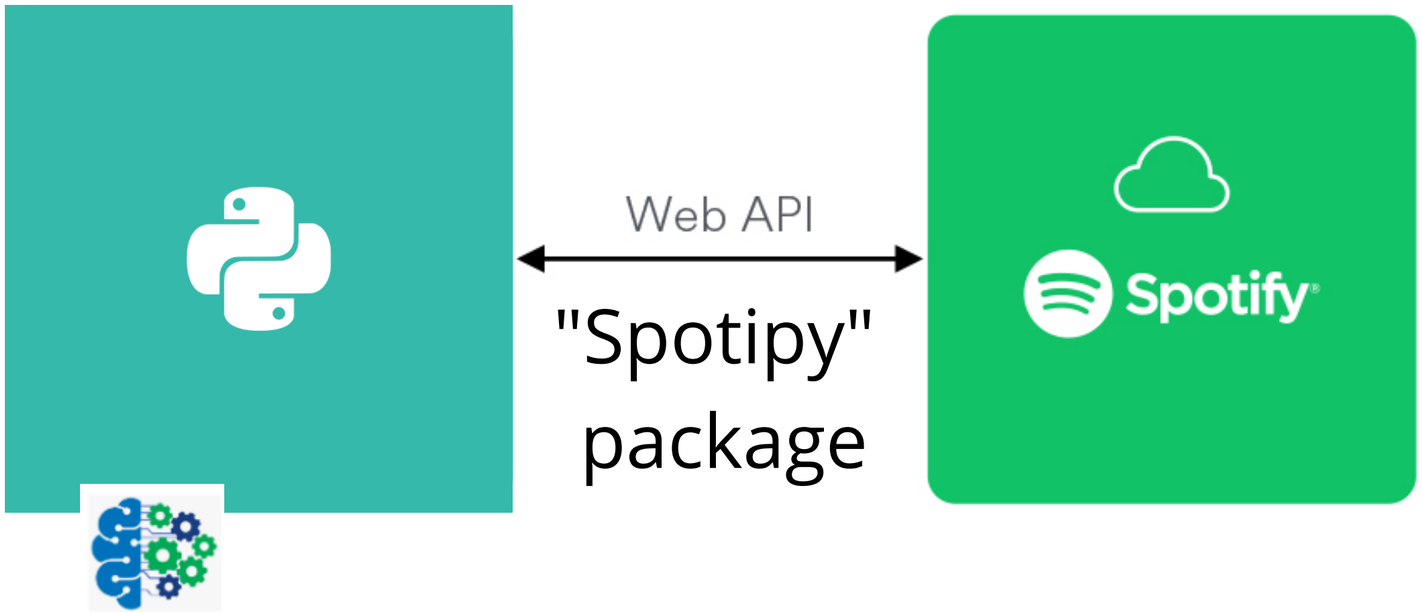
Part 2

Austin Heath

STAT 426
Data Science Methods
Final Project

December 10, 2020

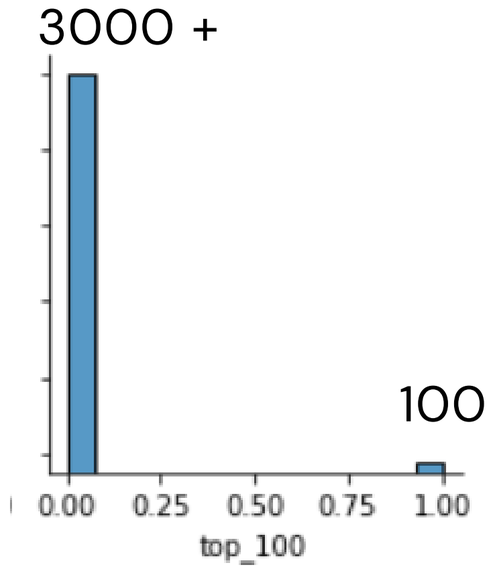
Overview

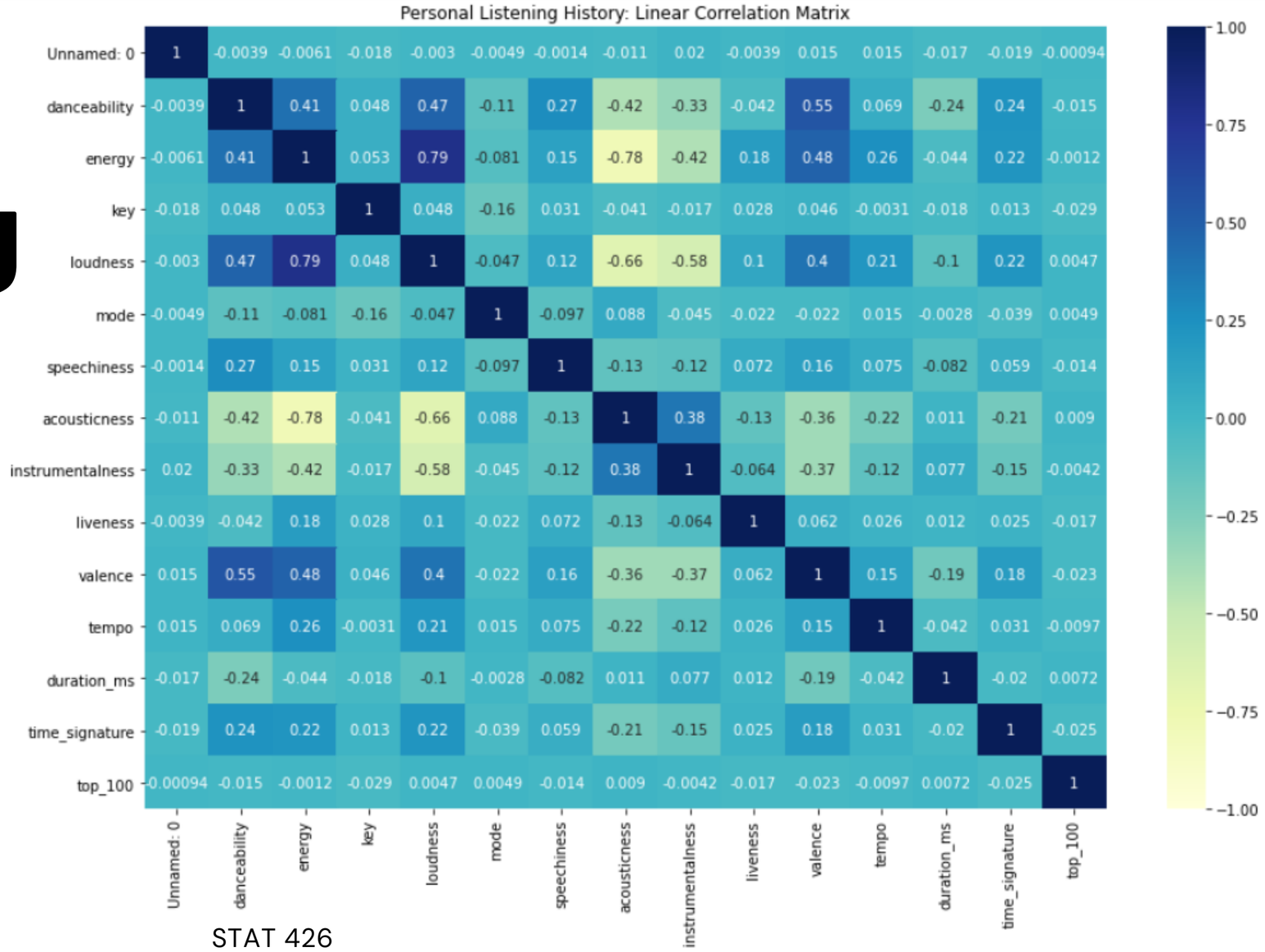


Random Forest Model
Gradient Boosting Model

Data Pre-Processing

Class Imbalances





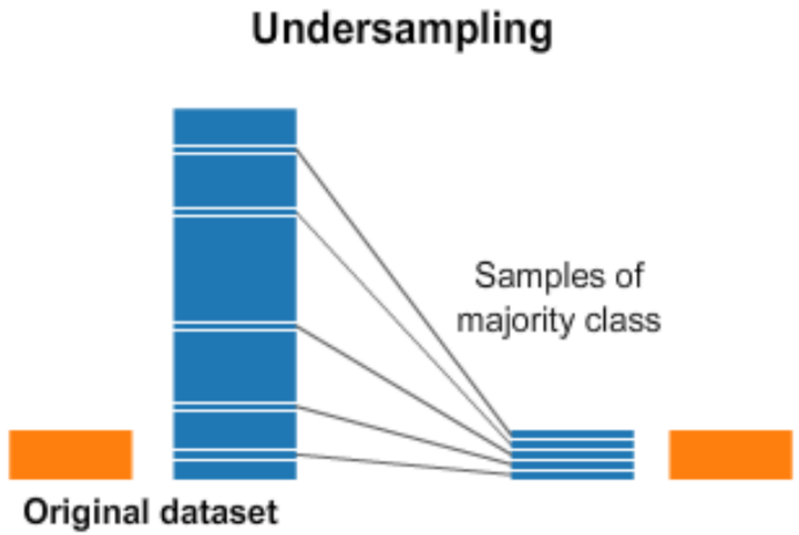
Austin Heath

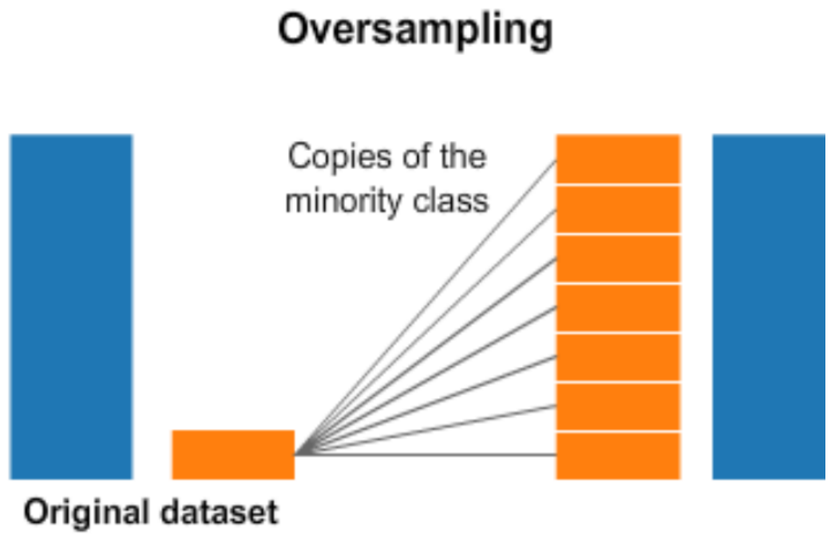
STAT 426
Data Science Methods
Final Project

December 10, 2020

Class Imbalances

SMOTE Package (Oversampling)
Near Miss Package (Undersampling)





Results

Part 3

Austin Heath

STAT 426
Data Science Methods
Final Project

December 10, 2020

Objectives

- Using training data from my personal Spotify listening history in 2020, can I accurately classify whether or not songs are in my top 100 songs of the year?
- Using this model, can I recommend new favorite songs?

```
SMOTE classification
accuracy: 0.9501100513573001
precision: 0.05
recall: 0.0625
f1: 0.05555555555555556

SMOTE Gradient Boost classification
accuracy: 0.8253851797505503
precision: 0.018691588785046728
recall: 0.125
f1: 0.03252032520325203

NearMiss classification
accuracy: 0.22670579603815114
precision: 0.019736842105263157
recall: 0.65625
f1: 0.0383211678832116 74

NearMiss Gradient Boost classification
accuracy: 0.2090975788701394
precision: 0.024545454545454544
recall: 0.84375
f1: 0.04770318021201413
```

Objectives

- Using training data from my personal Spotify listening history in 2020, can I accurately classify whether or not songs are in my top 100 songs of the year?
- Using this model, can I recommend new favorite songs?

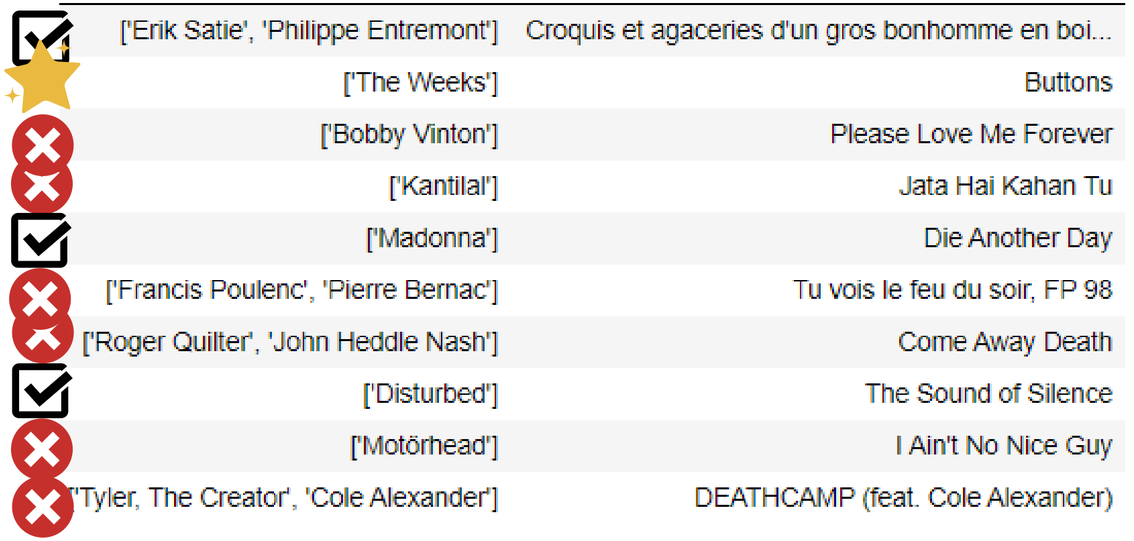
- All of the models performed poorly
- My top 100 songs represent diverse audio characteristics. According to the current variables in my training data, I conclude that I don't have a music type.

Next Steps:

Could including additional datapoints (such as popularity rating, release date or genre) improve model performance?

Objectives

- Using training data from my personal Spotify listening history in 2020, can I accurately classify whether or not songs are in my top 100 songs of the year?
- Using this model, can I recommend new favorite songs?



	artists	name
☑	['Erik Satie', 'Philippe Entremont']	Croquis et agaceries d'un gros bonhomme en boi...
✗	['The Weeks']	Buttons
✗	['Bobby Vinton']	Please Love Me Forever
☑	['Madonna']	Die Another Day
✗	['Francis Poulenc', 'Pierre Bernac']	Tu vois le feu du soir, FP 98
✗	['Roger Quilter', 'John Heddle Nash']	Come Away Death
☑	['Disturbed']	The Sound of Silence
✗	['Motörhead']	I Ain't No Nice Guy
✗	['Tyler, The Creator', 'Cole Alexander']	DEATHCAMP (feat. Cole Alexander)