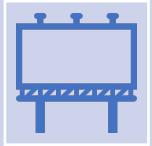


The background of the slide is a solid light purple color. Scattered around the central white box are several vintage cassette tapes. The tapes have various colored labels: blue, red, green, yellow, purple, and teal. Some labels have the letters 'A' or 'B' on them. The tapes are arranged in a circular pattern around the central text.

Predicting Top 10 Songs on Billboard Hot 100 Chart

Cassie Buhler

Billboard Charts



Billboard publishes musical charts of the most popular (highest sales, streams, etc.) songs & artists



Charts are released on a weekly and/or yearly basis













Great way to look at how trends and pop culture changes and evolves

The Hot 100

- *The Hot 100* is a popular Billboard chart
- Publishes 100 songs at end of year
- Songs are listed in ascending order– most popular at #1

Top 10 songs in 2010's Hot 100 Chart

1		Tik Tok Ke\$ha	▼
2		Need You Now Lady Antebellum	▼
3		Hey, Soul Sister Train	▼
4		California Gurls Katy Perry Featuring Snoop Dogg	▼
5		OMG Usher Featuring will.i.am	▼
6		Airplanes B.o.B Featuring Hayley Williams	▼
7		Love The Way You Lie Eminem Featuring Rihanna	▼
8		Bad Romance Lady Gaga	▼
9		Dynamite Taio Cruz	▼
10		Break Your Heart Taio Cruz Featuring Ludacris	▼

The Dataset

Hot 100 songs from 2010-2020 yearly charts

- **1100 songs** in total

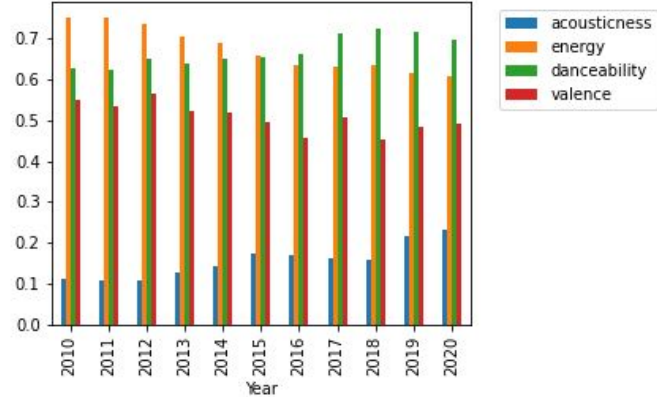
Split data into 3 sets

- Train: 2010-2015
- Validate: 2016-2017
- Test: 2018-2020

Spotify API Features:

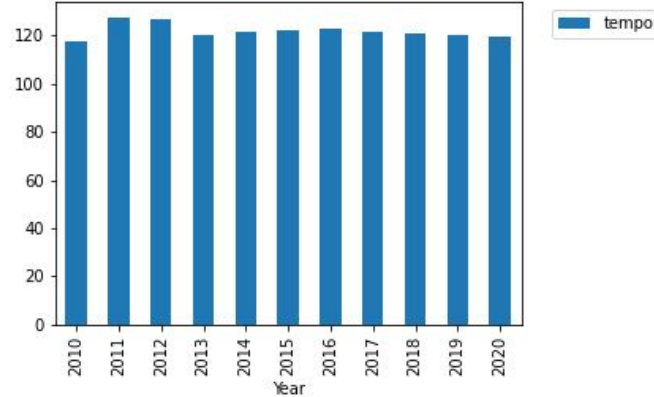
- Acousticness
- Danceability
- Duration
- Energy
- Instrumentalness
- Key
- Liveness
- Loudness
- Speechiness
- Tempo
- Time signature

Average Audio Features of Top 100 Songs: 2010-2020



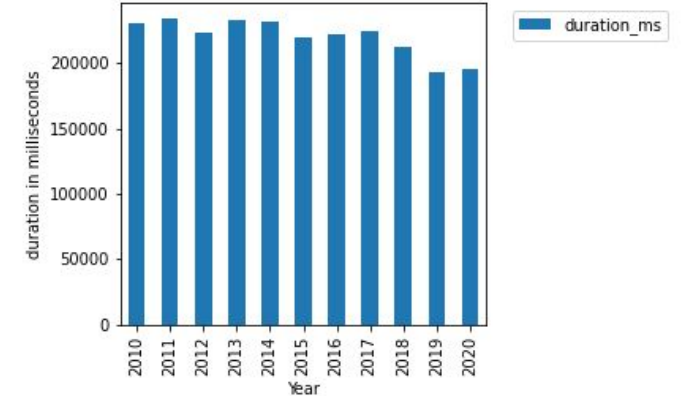
- Increase in acousticness & danceability
- Decrease in energy

Average Tempo of Top 100 Songs: 2010-2020



- Tempo is consistent

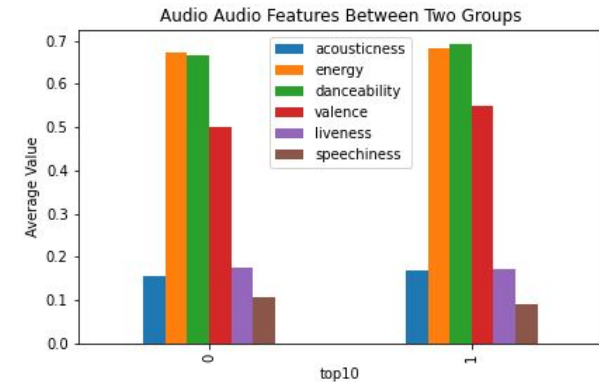
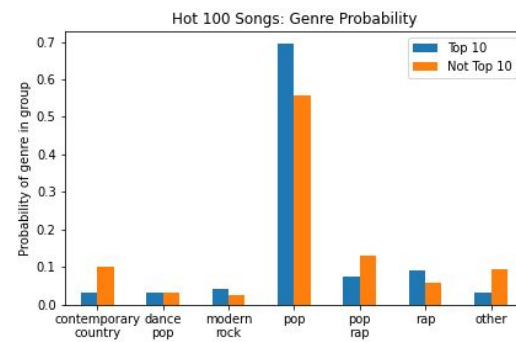
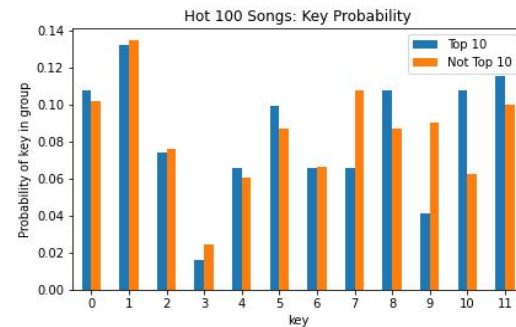
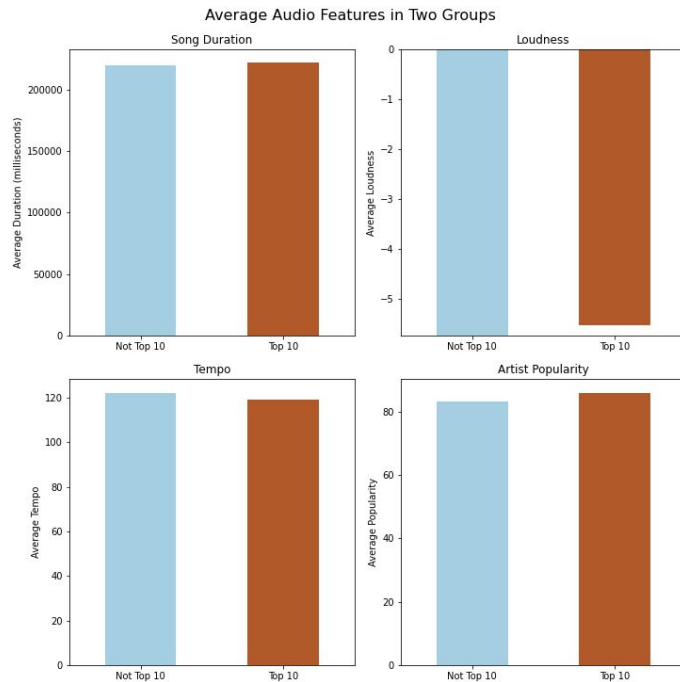
Average Length of Top 100 Songs: 2010-2020



- Decrease in song length

How has the Hot 100 changed over the past decade?

How do **Top 10** songs compare to the rest of the Hot 100?



- Higher valence and danceability
- Less loud, slower tempo
- More popular artists
- Pop and rap are most common



The Project

Which **features** best predict the **top 10** songs on Billboard Hot 100 charts?

Project Use & Intentions:

- Artists wanting to reach larger audiences can know what to focus on
- Music companies seek out songs that meet these features
- Predicting trends

Project Overview

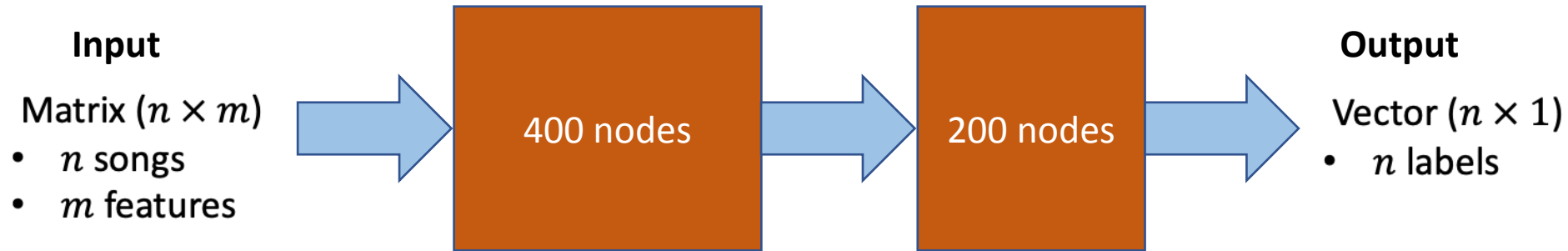
Use neural networks to solve binary classification problem

- Class 0: Not top 10 song (90% of dataset)
- Class 1: Top 10 song (10% of dataset)

Reduce dataset by using **feature selection** techniques

- Irrelevant features can impact the accuracy of our model
- Which features yield the best predictive model?

Classification with Neural Networks



1. Train network with training data
2. Use validation data to improve parameters
3. Feed network testing data
4. Use the network's output and compare with true classes of testing data

Applying Feature Selection to Neural Networks

Train and test the neural network using data that contains:

All features

Features selected by
ExtraTreesClassifier

Features selected by
Lasso

Features selected by
SVM

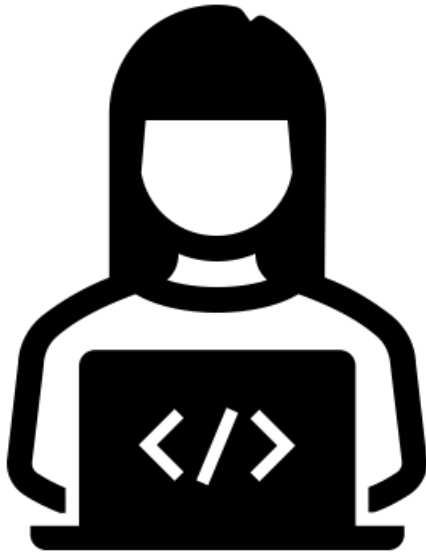


The model with the highest accuracy contains the features most useful with identifying top 10 songs

Results

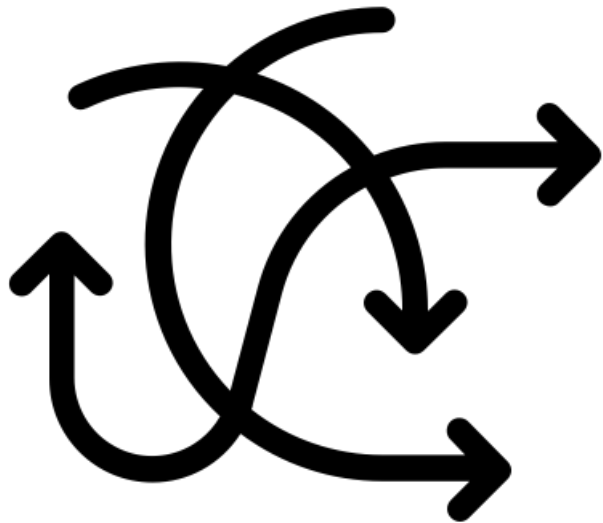
Method	ExtraTreesClassifier	SVM	Lasso	No Feature Selection
	<p>Feature Importance using ExtraTreesClassifier Model</p>	<p>Selecting Features using SVM Model</p>	<p>Selecting Features using Lasso Model</p>	
# Features	10	10	5	14
	<p>Confusion Matrix</p>	<p>Confusion Matrix</p>	<p>Confusion Matrix</p>	<p>Confusion Matrix</p>
Accuracy	74%	75%	55%	70%

The Team



- I worked alone on this project
- Skills: programming, machine learning, listens to music
- I collected the data and ran the analysis

Project Challenges & Limitations



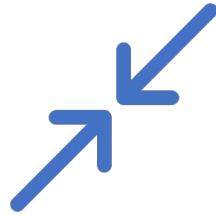
- Spotify API was irritating
 - Features are only for songs in a playlist or on the same album
 - Had to manually create Spotify playlists for 1100 Billboard songs
- Limited Data
 - Unable to obtain artists net worth demographics
 - No database for this information

Conclusion



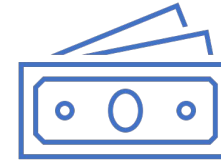
SVM performed the best with the following features:

Artist popularity
Danceability
Valence
Genre
Acousticness
Energy
Tempo
Instrumentalness
Key
Speechiness



Lasso performed the worst

Had smallest # of features
Likely too much data reduction



It'd be interesting to consider additional features

artist net worth
artist demographics