

What Makes a Hit Song?

Joe Swing, Lorin Helfenstein, Colette Crowder



Summary

- Goal
 - What factors make songs popular?
 - Lead to reaching the top 100 position charts
- Create models
- Generate Predictions

Outline

- Business Problem
- Data
- Methods
- Results
- Conclusions



Business Problem

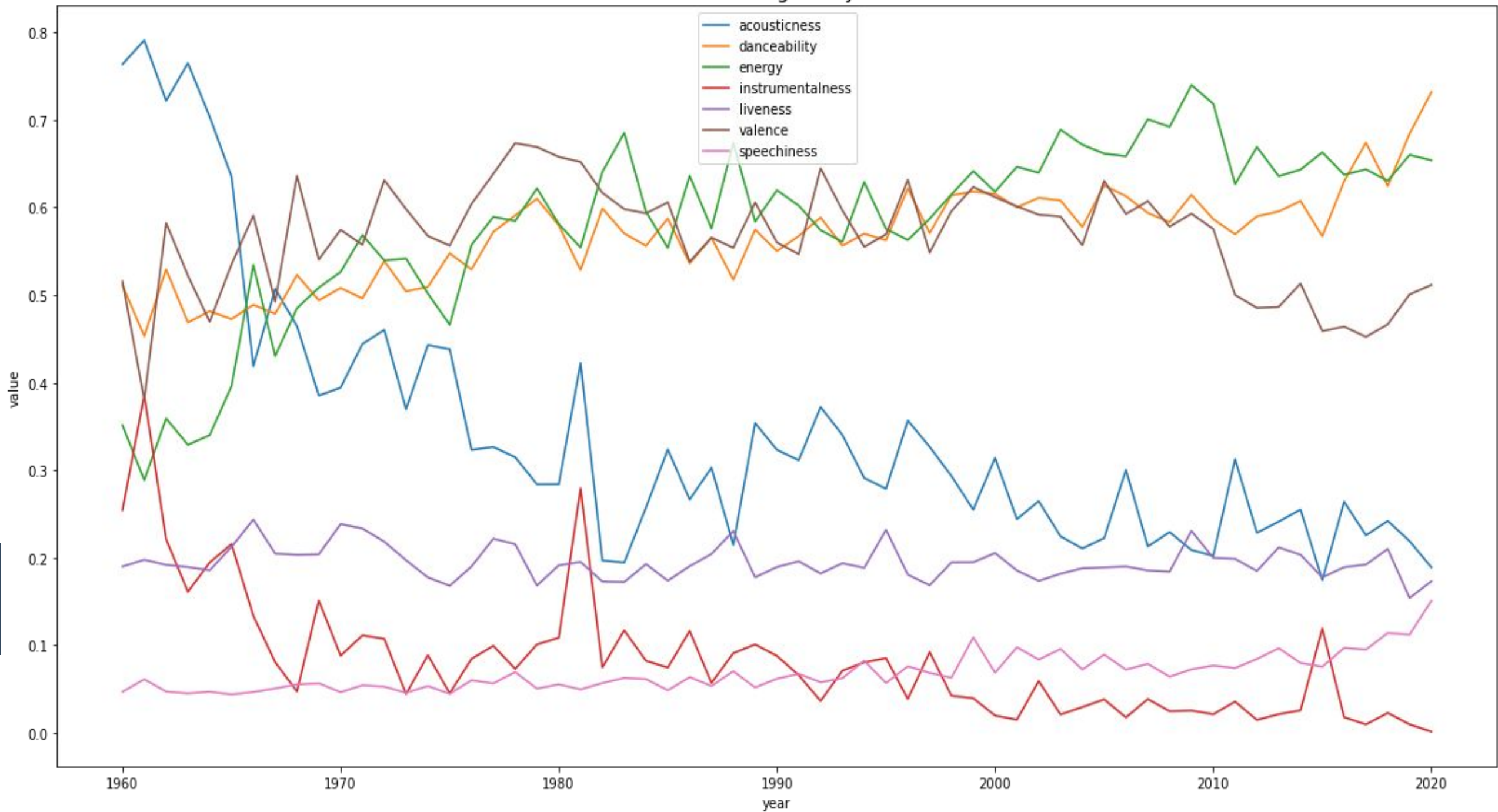
- A pop musical artist approaches us
- What kind of song will be a hit?
- What factors lead to popularity?
- Recommendations

Data

- Two datasets of songs
 - Overlap
 - Is it on the top 100?
 - Popularity
- Predictor Variables
 - ex. loudness, danceability, speechiness
- Datasets from Kaggle
 - Webscraping from billboard, shazam
- Spotify Web API



Trend of Song over years



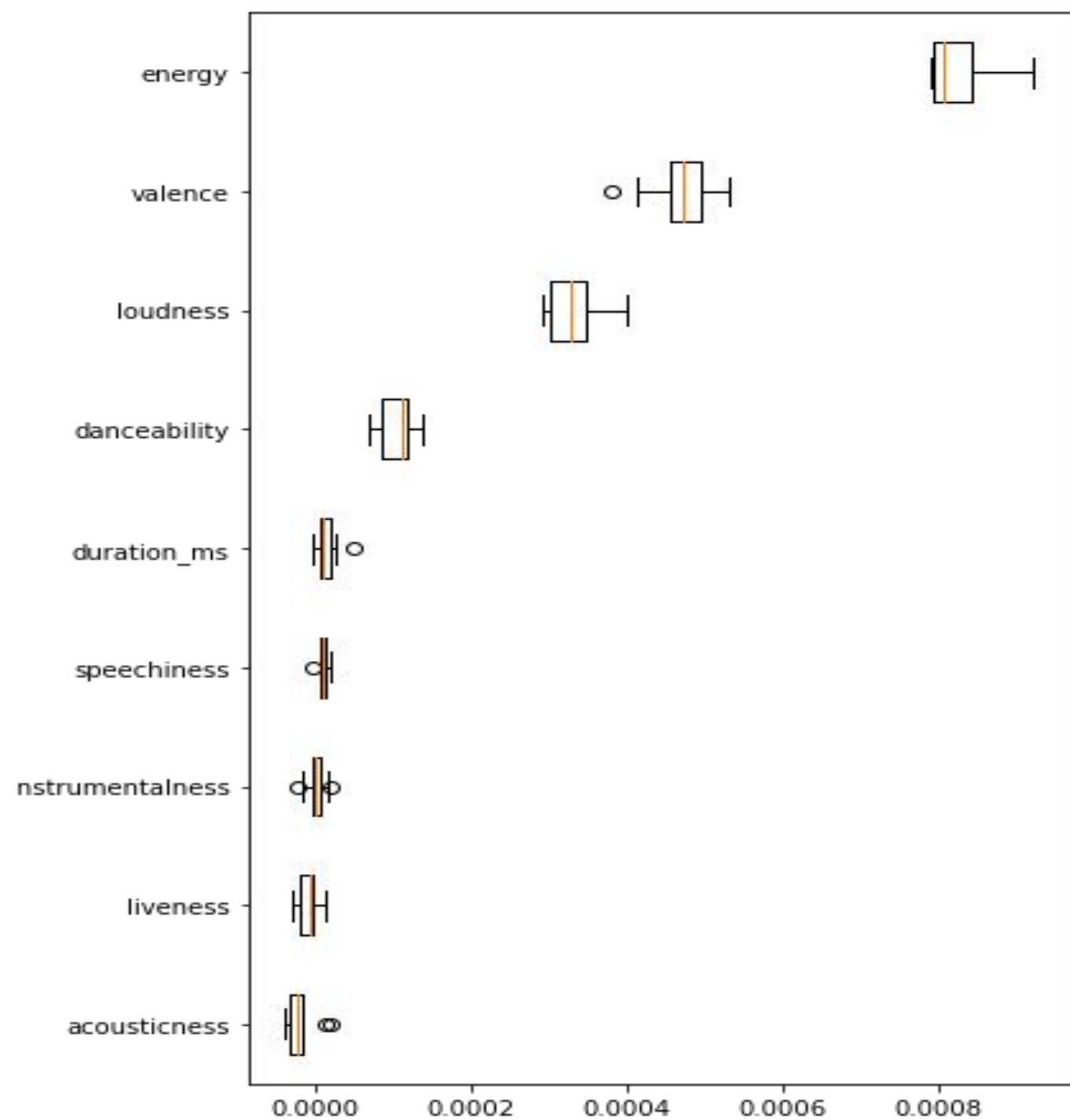
Methods

- In this project, we made and used multiple different models including:
 - Neural Networks
 - Decision Tree
 - Bagged Tree
 - Random Forrest
 - XGBoost
 - Logistic Regression
 - Grid Searching

Target

- The target variable is popularity
- I decided to go with the baseline logistic regression model over the neural network because both models gave similar scores
- Our model was able to get a validation accuracy of .926, or 92.6%.

Permutation Importance for Logistic Regression Model



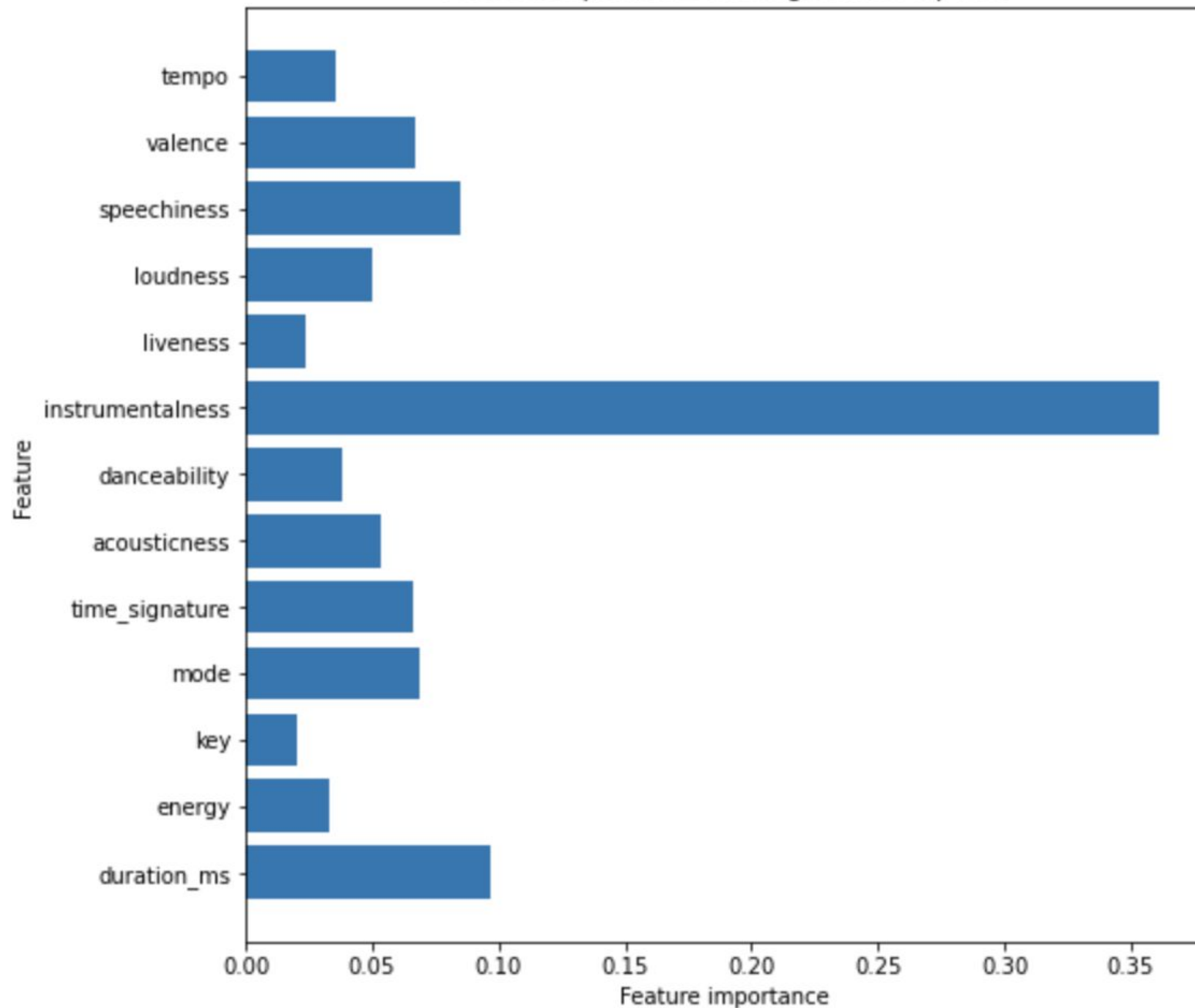
Popular Song Features

- Median loudness of a popular song is -6.5
 - Scale from -60 to 0
 - Popular songs tend to be loud
 - The attribute of a sound that determines the magnitude of the auditory sensation produced
- Median song energy is .662
 - Energy ranges from 0 to 1
 - Energy represents a perceptual measure of intensity and activity
- Median valence is .53
 - Valence is musical positivity
 - Higher valence = happier sounding
Lower valence = sad/angry sounding
 - On a scale from 0 to 1, 1 being the highest

Target and accuracy score

- The target variable is whether or not a song is on the top 100 hits or not
- Using a Grid Search on an XGBooster got us the best results, with an accuracy score of 81.26%.

Feature Importance of Songs in the Top 100



Conclusions

- If we had more time, we would have continued to work on the Spotify API and gotten the release dates for the songs to filter them and only use those songs from later than 2010.

Thank You!

Joe Swing

Email: jcswing@bsc.edu

Github: [@Jswing450](https://github.com/Jswing450)

Lorin Helfenstein

Email: lehelfen@bsc.edu

Github: [@lorinhelfenstein](https://github.com/lorinhelfenstein)

Colette Crowder

Email: crcrowde@bsc.edu

Github: [@crcrowde](https://github.com/crcrowde)

