MACHINE LEARNING FOR MUSIC: PREDICTING USER ENGAGEMENT USING SONG CHARACTERISTICS

API 222 - Machine Learning & Big Data Analytics Project, April 2023

Rohini Iyer and Pranav Bhargava

LISTEN TO THE SONG AND GUESS ITS POPULARITY

POLLEV.COM/PRANAVB384

PROBLEM STATEMENT

We use ML techniques to predict a song's popularity/engagement based on the song's characteristics

Helps music studios and artists predict which songs are more likely to garner greater engagement

ML models can identify patterns and trends in data, making the talent selection process for music companies more data driven.

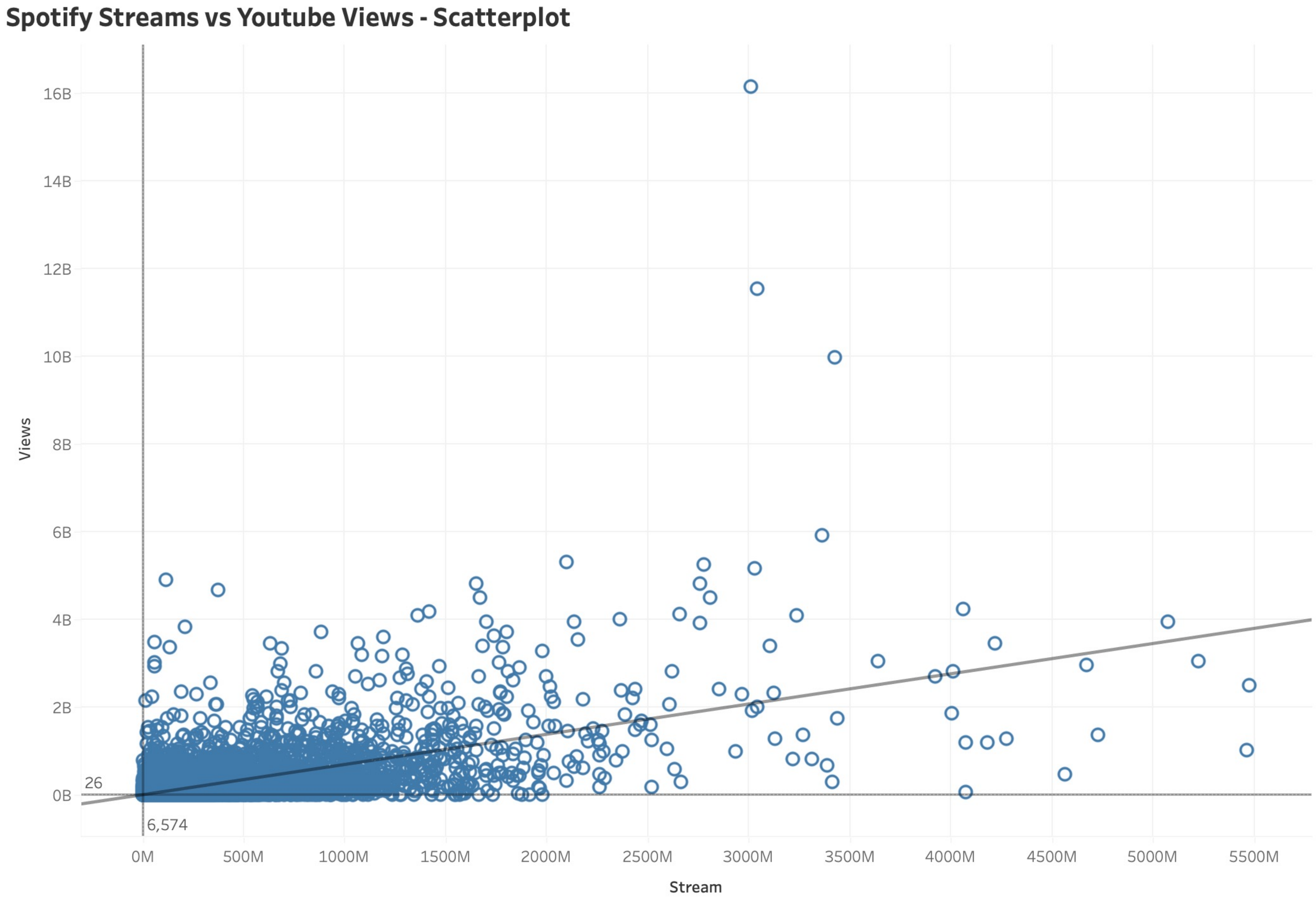
APPROACH

*Predict song popularity – **Streams per Views** based on song characteristics*

*Song characteristics / predictors like **Danceability, Key, Energy, Loudness, Speechiness, Acousticness, Instrumentalness, Liveness, Valence, Tempo***

We use multiple ML techniques to make this prediction

Songs with higher streams on Spotify also have higher views on Youtube



Correlation = 0.64

Almost 75% of songs have a higher number of Spotify Streams than Youtube Views

Summary of the outcome variable – **Streams Per Views (N = ~20k)**

Minimum	1st Quartile	Median (2nd Quartile	3rd Quartile	Maximum
0	1	3	12	38,637,558

Predictor	Definition
Danceability	**Describes how suitable a track is for dancing** based on a combination of musical elements including tempo, rhythm stability, beat strength, and overall regularity. A value of 0.0 is least danceable and 1.0 is most danceable.
Energy	Is a measure from 0.0 to 1.0 and represents a **perceptual measure of intensity and activity**. Typically, energetic tracks feel fast, loud, and noisy. For example, death metal has high energy, while a Bach prelude scores low on the scale. Perceptual features contributing to this attribute include dynamic range, perceived loudness, timbre, onset rate, and general entropy.
Key	**The key the track is in.** Integers map to pitches using standard Pitch Class notation. E.g. 0 = C, 1 = C♯/D♭, 2 = D, and so on. If no key was detected, the value is -1.
Loudness	The overall loudness of a track in decibels (dB). Loudness is the quality of a sound that is the **primary psychological correlate of physical strength** (amplitude). Values typically range between -60 and 0 db.
Speechiness	Detects **the presence of spoken words** in a track. The more exclusively speech-like the recording (e.g. talk show, audio book, poetry), the closer to 1.0 the attribute value. Values above 0.66 describe tracks that are probably made entirely of spoken words. Values between 0.33 and 0.66 describe tracks that may contain both music and speech, either in sections or layered, including such cases as rap music. Values below 0.33 most likely represent music and other non-speech-like tracks.
Acousticness	A **confidence measure** from 0.0 to 1.0 of whether the track is acoustic. 1.0 represents high confidence the track is acoustic.
Instrumentalness	Predicts **whether a track contains no vocals**. "Ooh" and "aah" sounds are treated as instrumental in this context. Rap or spoken word tracks are clearly "vocal". The closer the instrumentalness value is to 1.0, the greater likelihood the track contains no vocal content. Values above 0.5 are intended to represent instrumental tracks, but confidence is higher as the value approaches 1.0.
Liveness	Detects the **presence of an audience** in the recording. Higher liveness values represent an increased probability that the track was performed live. A value above 0.8 provides strong likelihood that the track is live.
Valence	A measure from 0.0 to 1.0 describing the **musical positiveness conveyed by a track**. Tracks with high valence sound more positive (e.g. happy, cheerful, euphoric), while tracks with low valence sound more negative (e.g. sad, depressed, angry).
Tempo	The overall **estimated tempo of a track in beats per minute (BPM)**. In musical terminology, tempo is the speed or pace of a given piece and derives directly from the average beat duration.
Views/Likes/Comments	Number of views/likes/Comments on Youtube
Streams	Number of Streams on Spotify

Source : Kaggle

MODELS SELECTED

- **Linear regression** – to establish a MSE baseline if there exists a linear relationship
- **KNN** – to take advantage of a non-parametric supervised learning model
- **Decision Tree** – easy to interpret and graph. Not computationally intensive (only 10 predictors)
- **Bagging** – further reduces variance of a decision tree while keeping the same bias
- **Random Forest** – provide an improvement over Bagging
- **Boosting** – another extension of decision trees that might give better performance

RESULTS

Mean Squared Error of Prediction (MSEP) :

Models explored	Streams/Views (Training) (in 10^8)	Streams/Views (Test) (in 10^8)
Ordinary Least Squares (OLS)	981.86	0.64
K-Nearest Neighbours (KNN)	984.76	2.61
Decision Tree	982.14	0.38
Random Forest	252.36	3.08
Bagging	494.63	3.04
Boosting	6.12	34.62

Predictors used: Danceability, Key, Energy, Loudness, Speechiness, Acousticness, Instrumentalness, Liveness, Valence, Tempo

Seed: 222 | **Original N:** 20718

Missing data: 1655 observations had missing data across one or multiple variables and hence were dropped from the data

Training data: Contains 80% of non-missing data observations = 15250 observations

Test data: Contains 20% of non-missing data observations = 3813 observations

REFLECTIONS AND NEXT STEPS

- **Analysis next steps**
  - Testing prediction accuracy
  - Classification of songs as "hits" or "misses" | assigning a popularity score
  - Deep diving into the predictors - are they correlated?

- **Limitations**
  - Missing song release data
  - Lack of user data

- **Concerns (with current data) that hinder adoption**
  - Calculating the values of the predictors
  - User engagement could also depend on factors other than song characteristics

THANK YOU!

APPENDIX

Results (v1)

Mean Squared Error of Prediction

Model fitted	Views (YouTube) (in 10^{16})	Likes (YouTube) (in 10^{12})	Streams (Spotify) (in 10^{16})
OLS	6.64	2.85	6.69
KNN	6.66 (k = 100)	2.86 (k = 187)	6.78 (k = 72)
Decision Tree	6.59	2.88	6.85
Random Forest	5.68	2.31	5.42
Bagging	5.85	2.37	5.50
Boosting	6.91	2.77	6.29

Predictors used: Danceability, Key, Energy, Loudness, Speechiness, Acousticness, Instrumentalness, Liveness, Valence, Tempo

Seed: 222 | **Original N:** 20718

Missing data: 1655 observations had missing data across one or multiple variables and hence were dropped from the data

Training data: Contains 80% of non-missing data observations = 15250 observations

Test data: Contains 20% of non-missing data observations = 3813 observations

Results (v2)

Mean Squared Error of Prediction (MSEP) :

Model fitted	Views/Streams (Training)	Views/Streams (Test)	Streams/Views (Training)	Streams/Views (Test)	Likes/Views (Training)	Likes/Views (Test)
OLS	12172.73	3860.424	98,186,250,736	64,008,591	0.0001229585	0.0001093745
KNN	12261.9	3944.848	98,475,885,939	261,292,981	0.0001296594	0.0001138033
Decision Tree	12177.8	3860.378	98,213,651,158	37,818,826	0.0001265753	0.0001122965
Random Forest	3050.763	3950.387	25,235,877,532	307,857,387	0.00002539191	0.0001027211
Bagging	3858.163	6241.07	49,462,853,002	303,533,254	0.00002380341	0.0001063668
Boosting	122.1899	4842.992	611,915,839	3,461,813,651	0.00004074487	0.0001249777

Predictors used: Danceability, Key, Energy, Loudness, Speechiness, Acousticness, Instrumentalness, Liveness, Valence, Tempo

Seed: 222 | **Original N:** 20718

Missing data: 1655 observations had missing data across one or multiple variables and hence were dropped from the data

Training data: Contains 80% of non-missing data observations = 15250 observations

Test data: Contains 20% of non-missing data observations = 3813 observations

Results (v3)

Mean Squared Error of Prediction (MSEP) :

Model fitted	Views/Streams (Training)	Views/Streams (Test)	Streams/Views (Training)	Streams/Views (Test)
OLS	12172.73	3860.424	98,186,250,736	64,008,591
KNN	12261.9	3944.848	98,475,885,939	261,292,981
Decision Tree	12177.8	3860.378	98,213,651,158	37,818,826
Random Forest	3050.763	3950.387	25,235,877,532	307,857,387
Bagging	3858.163	6241.07	49,462,853,002	303,533,254
Boosting	122.1899	4842.992	611,915,839	3,461,813,651

Predictors used: Danceability, Key, Energy, Loudness, Speechiness, Acousticness, Instrumentalness, Liveness, Valence, Tempo

Seed: 222 | **Original N:** 20718

Missing data: 1655 observations had missing data across one or multiple variables and hence were dropped from the data

Training data: Contains 80% of non-missing data observations = 15250 observations

Test data: Contains 20% of non-missing data observations = 3813 observations

Path forward

- **Scope (open questions/things to do)**
  - Classification of upcoming songs as potential "hits" or "misses"
  - Grouping songs into "genres" or "categories" to improve user experience
  - Deep diving into the predictor - are the different predictors correlated?

- **Limitations**
  - Inability to predict hits due to missing song release data
  - Inability to do recommendation analysis due to lack of user data

- **Concerns (with current data) that hinder adoption**
  - Calculating the values of the predictors
  - User engagement could also depend on factors other than song characteristics