Machine Learning for Music Genre Classification

Random Forest & Gradient Boosted Trees

Kevin Giroux

January 2021

Problem statement

Shazam is a company that matches user-recorded audio clips against their proprietary music database to identify the title and artist of any given track.

They are expanding their product to create a feature that will identify the genre of any track that their app is unable to identify.

As part of this effort, I have been hired to build out a proof of concept using track metadata obtained from Spotify.

Primary goals:

1. Develop an understanding for which features are most important to the musical genre classification problem and
2. Get a sense for whether genre classification based purely on an audio waveform or audio fingerprint might be possible.

Dataset overview

- For this proof of concept, Spotify data was used to determine whether or not characteristics of given track’s waveform / audio fingerprint can be used to predict the track’s genre

- The following features were included in the track metadata:

  - **Acousticness**
  - **Danceability**
  - **Energy**
  - **Instrumentalness**
  - **Liveness**
  - **Loudness**
  - **Speechiness**
  - **Valence**
  - **Tempo**
  - **Popularity**

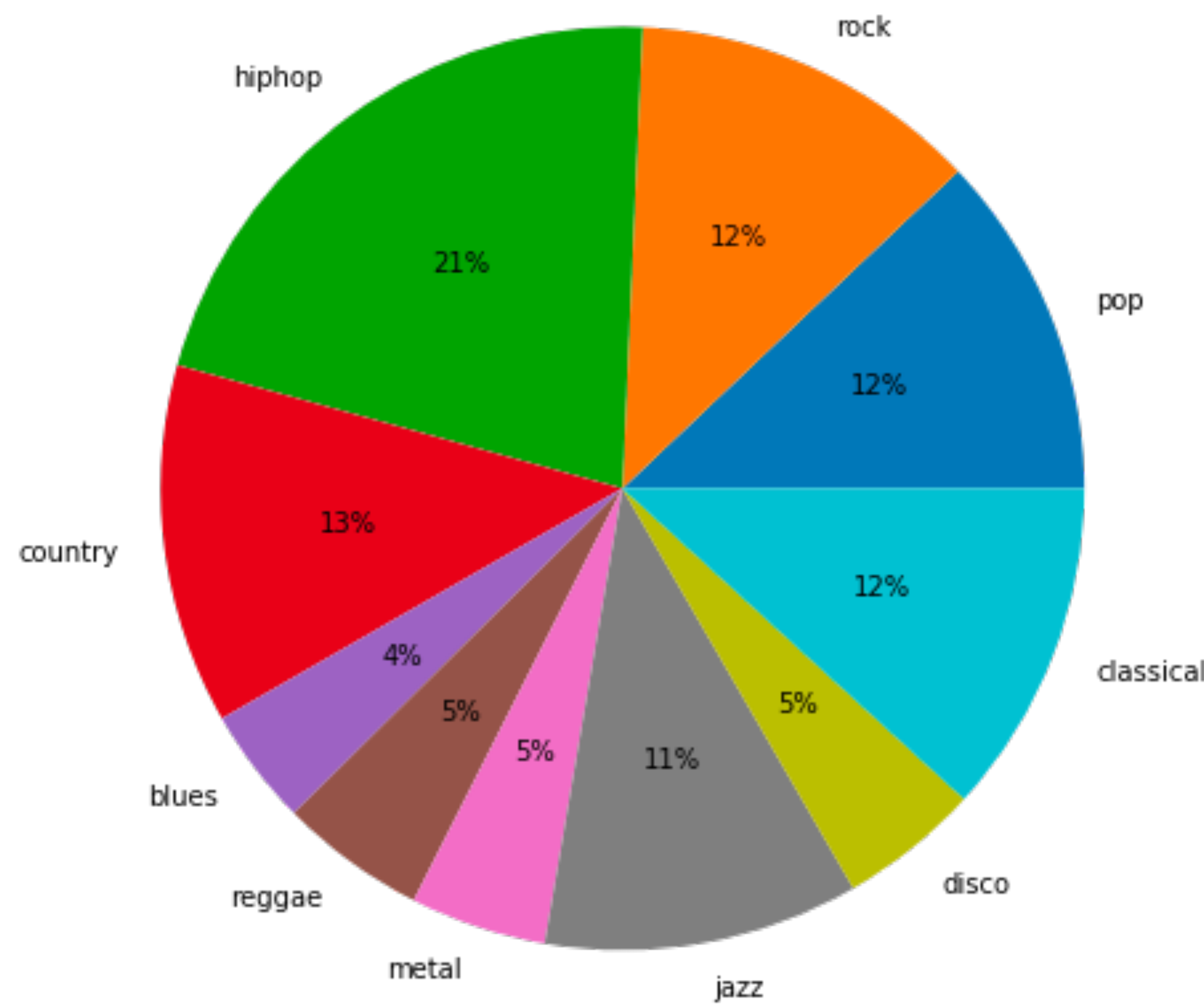
Predictive feature detail

- **Acousticness** - a confidence measure from 0.0 to 1.0 of whether the track is acoustic.
- **Danceability** - a measure of how suitable a track is for dancing based on a combination of musical elements including tempo, rhythm stability, beat strength, and overall regularity.
- **Energy** - 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.
- **Instrumentalness** - predicts whether a track contains no vocals.
- **Liveness** - a confidence measure that detects the presence of an audience in the recording.
- **Loudness** - the overall loudness of a track in decibels (dB), as averaged across the entire track
- **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.
- **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** - a measure of the speed of a track, in beats per minute.
- **Popularity**

Target variable: Genre

Total sample size: 2,438

- Pop (287)
- Rock (301)
- Hip hop (546)
- Country (308)
- Blues (101)
- Reggae (120)
- Metal (114)
- Jazz (265)
- Disco (115)
- Classical (281)



Methodology

- Defined macro-genre categories for prediction; manually grouped Spotify track data into macro-genre buckets based on sub-genre tagging at the artist level

- Trained 9 different classification algorithms to identify the algorithms

- Fine-tuned top-performing algorithms to optimize for test prediction accuracy

- Used the resulting model to answer the following questions:

Questions for analysis:

- Is Spotify track metadata useful as a predictor of track genre?
- What genres is the model best at identifying?
- Which genre predictions that come out of the model are the most certain?
- Which metadata features are the most important in genre prediction?
- Are genres distinguishable from each other based on average values for each feature?
- What should next steps be?

RESULTS

1. Is track metadata useful as a predictor of track genre?

- YES!
- All of the tested algorithms performed with at least 50% accuracy, which is a huge improvement on random chance (10%)
- The best algorithms breached 65% accuracy after model parameters were fine-tuned
- NOTE: all the models were better-suited for predicting certain genres than others; in other words not every prediction is created equally

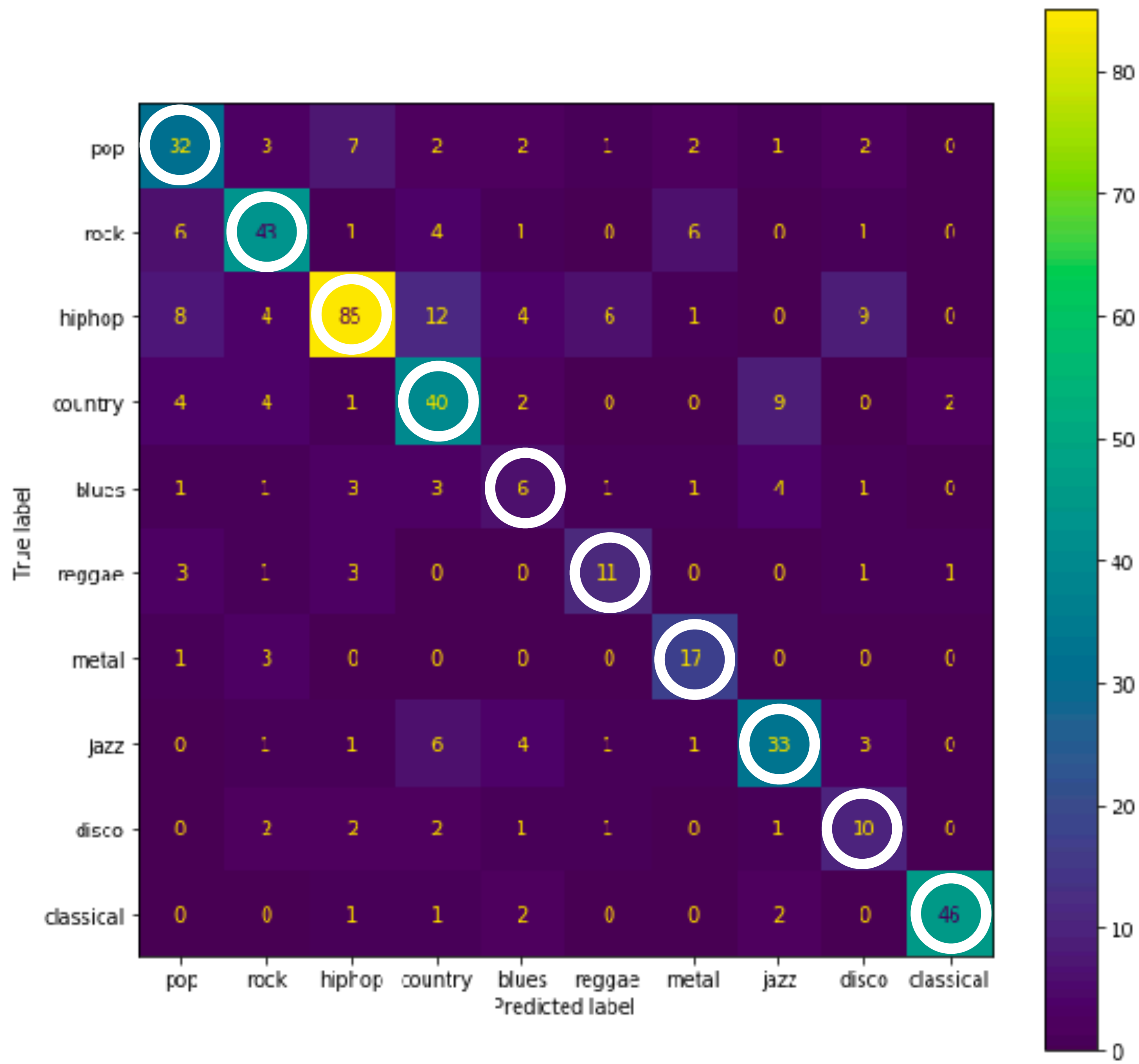
Recommendations:

- Proof of concept was a success; stage 2 of product development can proceed
- Differences in prediction accuracy between genres will inform next steps on the technical side

1. Gradient Boosted Trees

Overall model accuracy: 66%

(the highest of all algorithms tested)



NOTE: Read chart diagonally downwards from left to right to see the total number of CORRECT genre predictions for each genre.

1. Gradient Boosted Trees

A. How good is the model at identifying each genre?

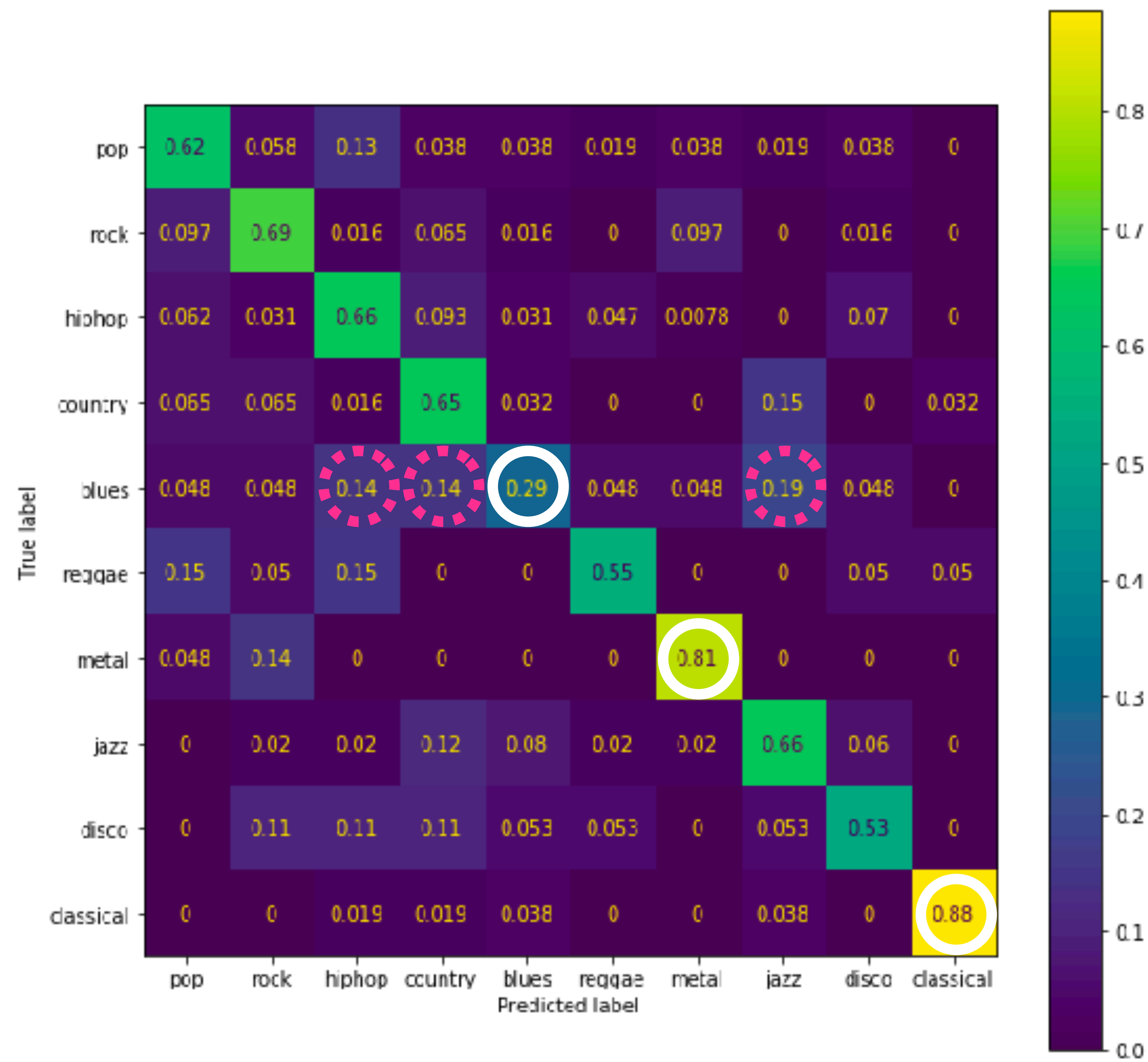
Most genres were correctly identified at a rate of between 58% and 66%.

Strengths

- 88% of all CLASSICAL tracks in the dataset were correctly identified
- 81% of all METAL tracks in the dataset were correctly identified

Weaknesses

- 47% of all BLUES tracks were incorrectly classified as either HIP HOP, COUNTRY, or JAZZ



% of BLUES tracks that were misclassified

1. Gradient Boosted Trees

B. How reliable is the model's genre prediction?

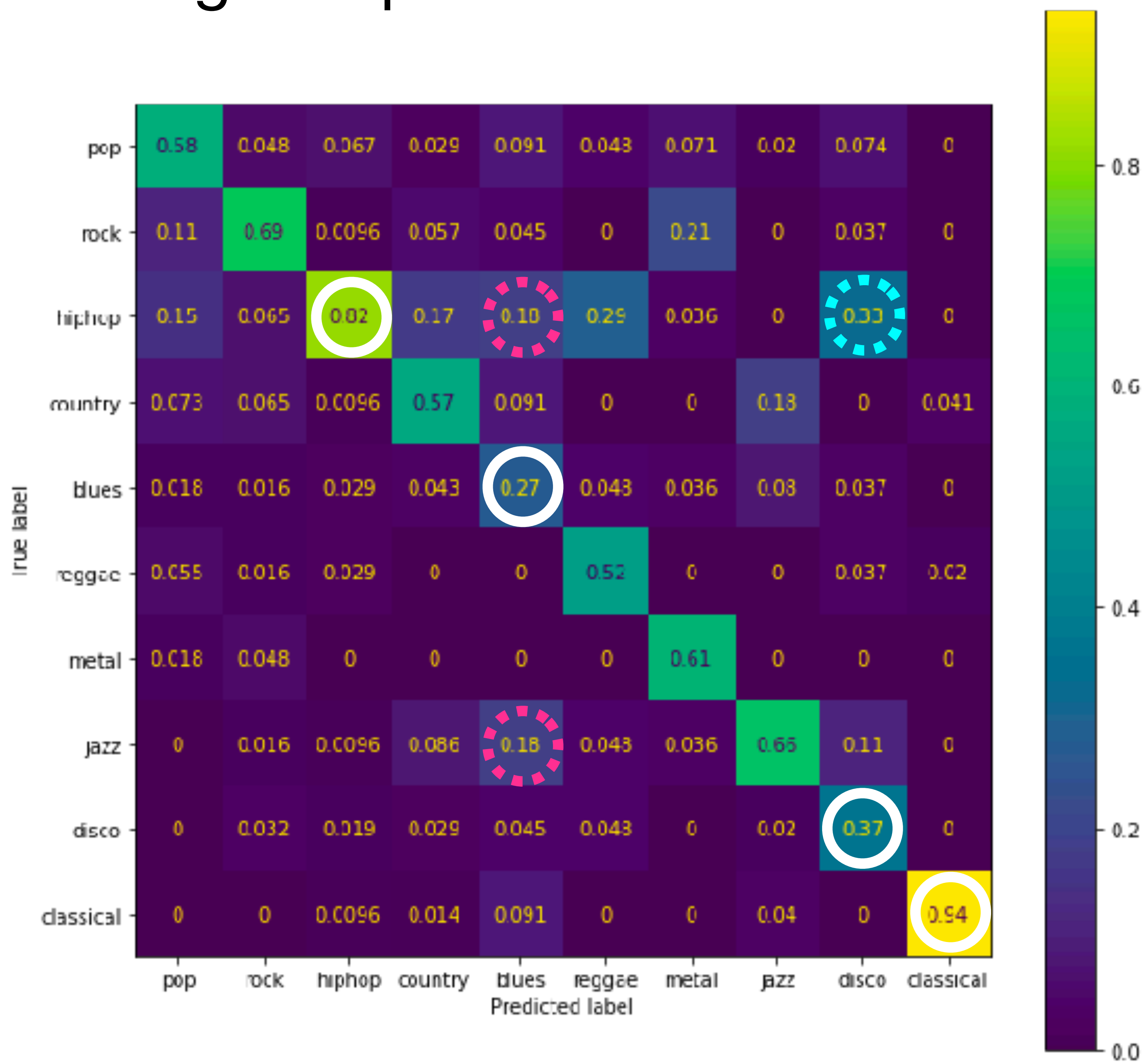
Most genre predictions are correct between 45% and 65% of the time.

Strengths

- 94% of all CLASSICAL predictions were correct
- 82% of all HIP HOP predictions were correct

Weaknesses

- 36% of all predicted BLUES tracks were actually either JAZZ or HIP HOP
- 33% of all predicted DISCO tracks were actually HIP HOP

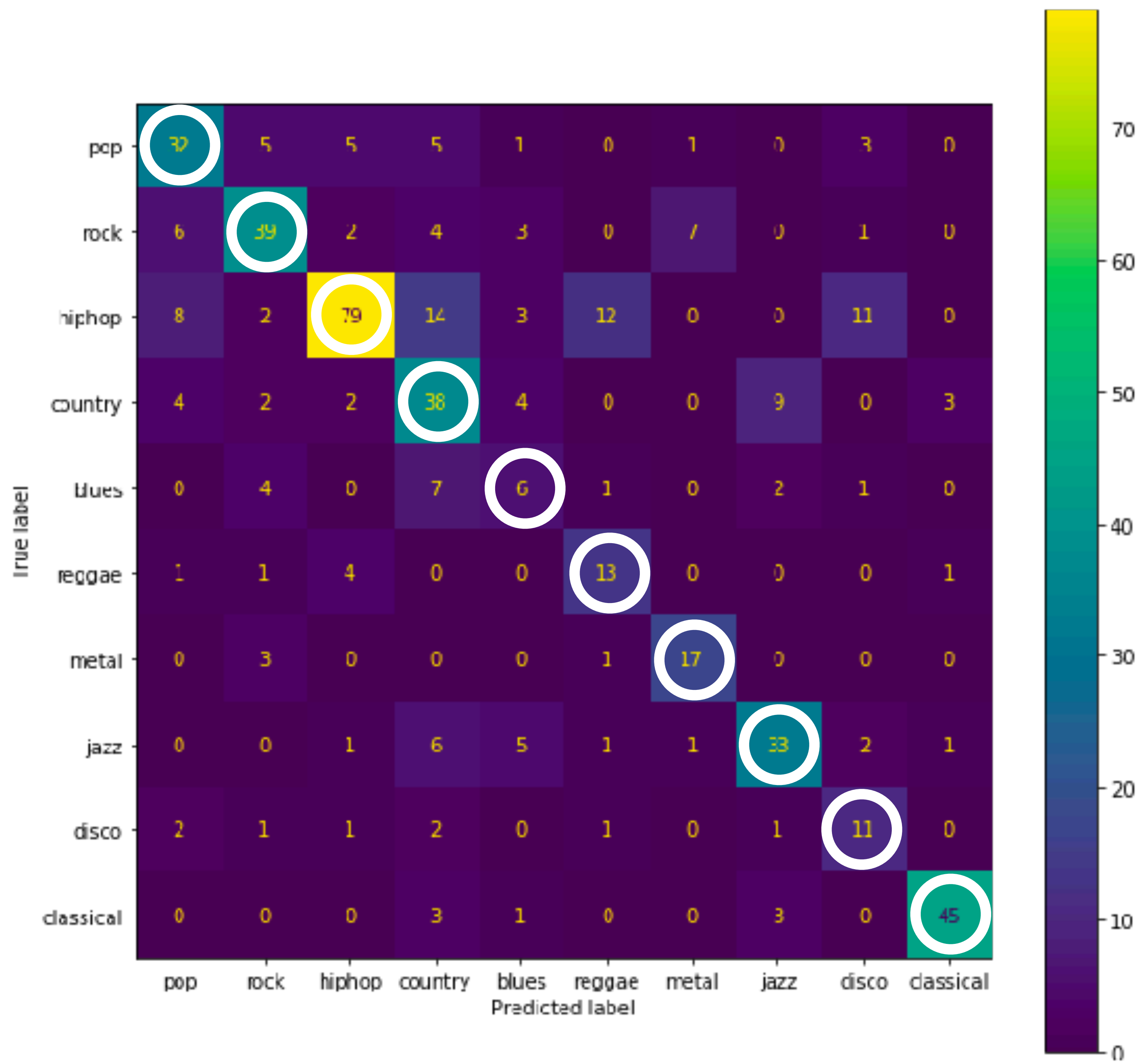


% incorrectly identified as BLUES   % incorrectly identified as DISCO

2. Random Forest

Overall model accuracy: 64%

(the second-highest of all algorithms tested)



NOTE: Read chart diagonally downwards from left to right to see the total number of CORRECT genre predictions for each genre.

2. Random Forest

A. How good is the model at identifying each genre?

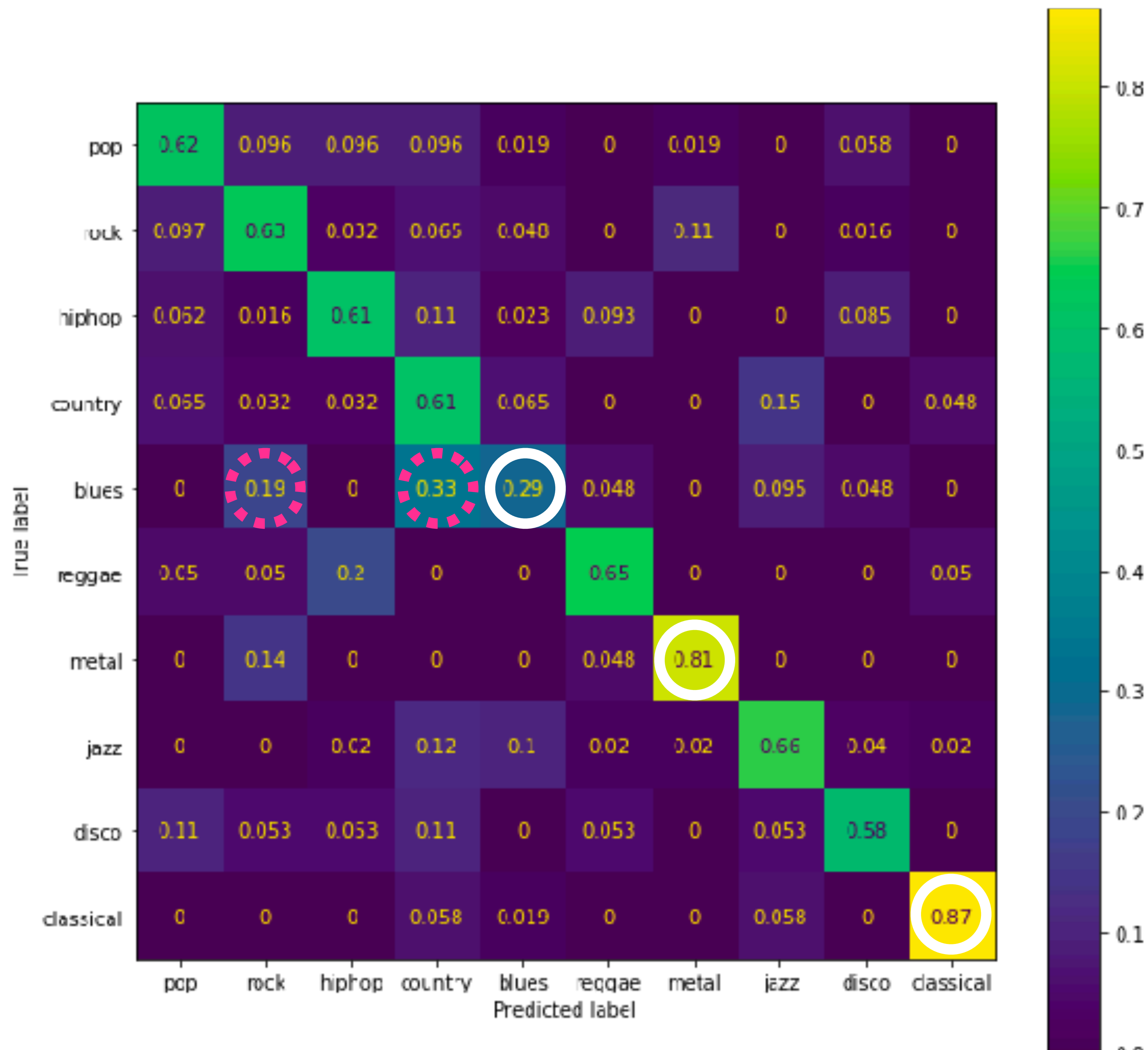
- Most genres were correctly identified with an accuracy of between 58% and 66%

Strengths

- 87% of all CLASSICAL tracks in the dataset were correctly identified
- 81% of all METAL tracks in the dataset were correctly identified

Weaknesses

- 52% of all BLUES tracks in the dataset were incorrectly classified as either ROCK or HIP HOP



% of BLUES tracks that were misclassified

2. Random Forest

B. How reliable is the model's genre prediction?

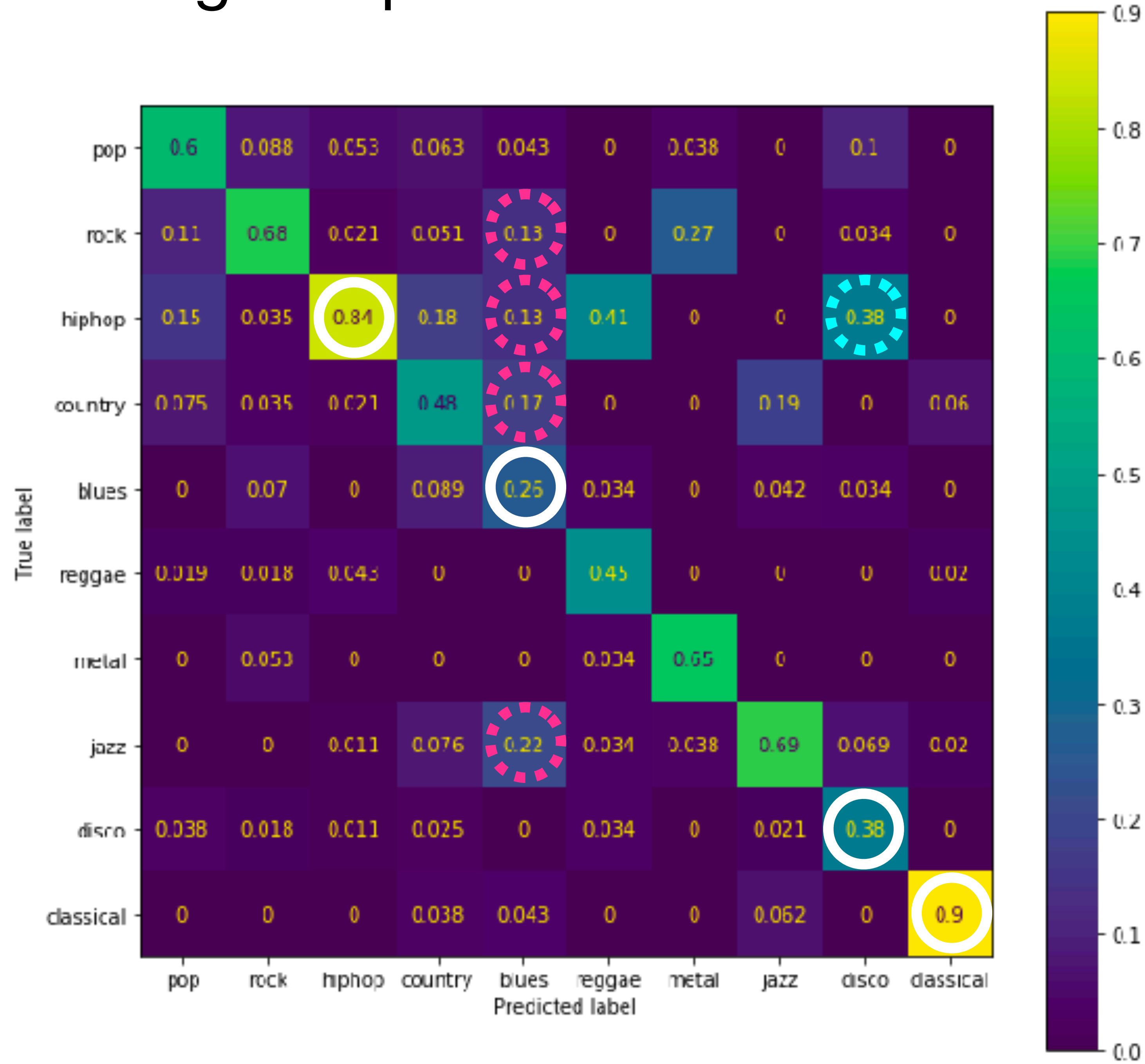
- Most genre predictions are correct between 45% and 65% of the time

Strengths

- 90% of all CLASSICAL predictions were correct
- 84% of all HIP HOP predictions were correct

Weaknesses

- 65% of all predicted BLUES tracks were actually JAZZ, COUNTRY, HIP HOP, or ROCK
- 38% of all predicted DISCO tracks were actually HIP HOP



% incorrectly identified as BLUES

% incorrectly identified as DISCO

What genres is the model best at identifying?

- Both the Gradient Boosted Trees and Random Forest algorithms were best at identifying CLASSICAL and METAL tracks

- Both models were worst at identifying BLUES tracks

What predicted genres are the most reliable?

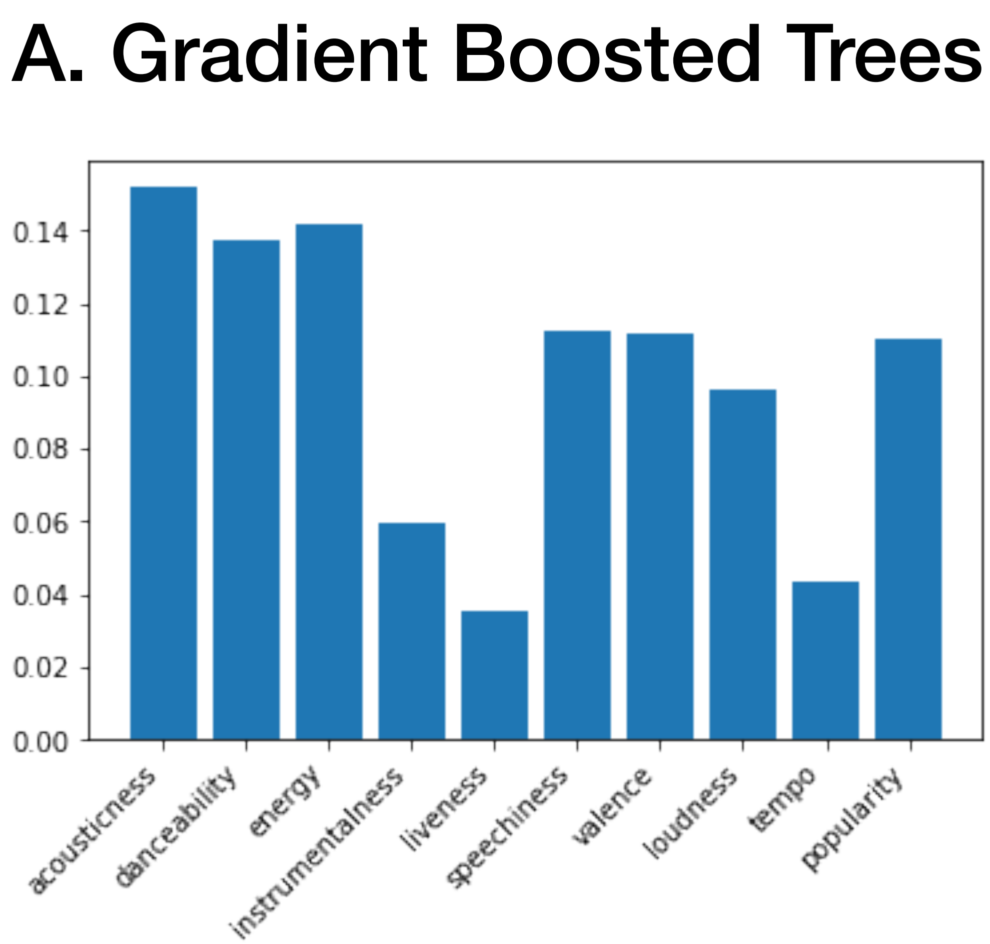
- Predictions of CLASSICAL and HIP HOP are the most reliable from both models

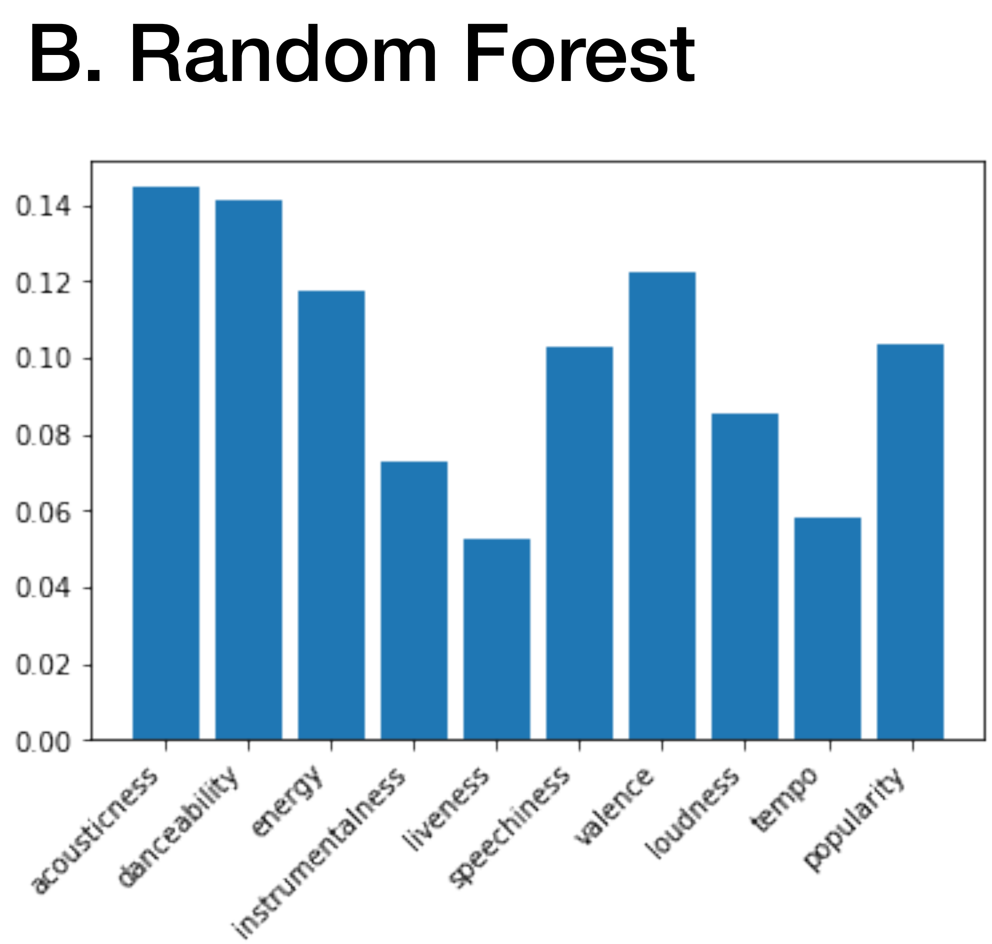
- Predictions of BLUES and DISCO are the least reliable from both models

Recommendations:

- The model, and therefore genre prediction products generally, work best on extreme-sounding genres
- Expand our dataset to include more unique examples of BLUES and DISCO
- Develop a feature that provides 2nd and 3rd choice predictions for harder-to-predict genres

Top important features





Top 6 important features (both models): Acousticness, Energy, Danceability, Speechiness, Valence, Popularity

__Recommendation__:

- Data derived directly from an audio waveform / fingerprint can likely be used for the task of genre classification, based on most important features
- However, one feature, Popularity, is not a derivative of the waveform, indicating metadata outside of the audio fingerprint may help to improve the accuracy of predictions based on an audio waveform

Business Recommendations

1. Based on the preceding analysis, it appears data derived from an audio waveform IS useful as a predictor for the track's genre; thus we should greenlight phase 2 of app development.

2. Though genre prediction via track waveform seems feasible, prediction accuracy may be improved with the addition of non-audio metadata features. This should be considered by both the business and technology teams.

3. The models in their current form tend to work best on extreme-sounding genres (e.g. Classical, Metal, and Hip Hop), but tends to confuse similar sounding genres (e.g. Blues, Jazz, and Rock). Thus the final product should have a feature that offers 2nd and 3rd choice genre classification predictions when a given prediction is below a certain threshold of certainty.

Future work

1. Expand the dataset to include additional unique examples of Blues and Disco tracks, which were the most difficult for the models to classify correctly.

2. Experiment with building a series of models that are trained only on the top 6 most important features identified in this phase of the project.

3. Kick off an in-depth genre mapping exercise which will inform the training of future, more sophisticated versions of these models

Thank you!

Kevin Giroux
kevinsgiroux@gmail.com