Ethan Straub

STAA 578

Zhao

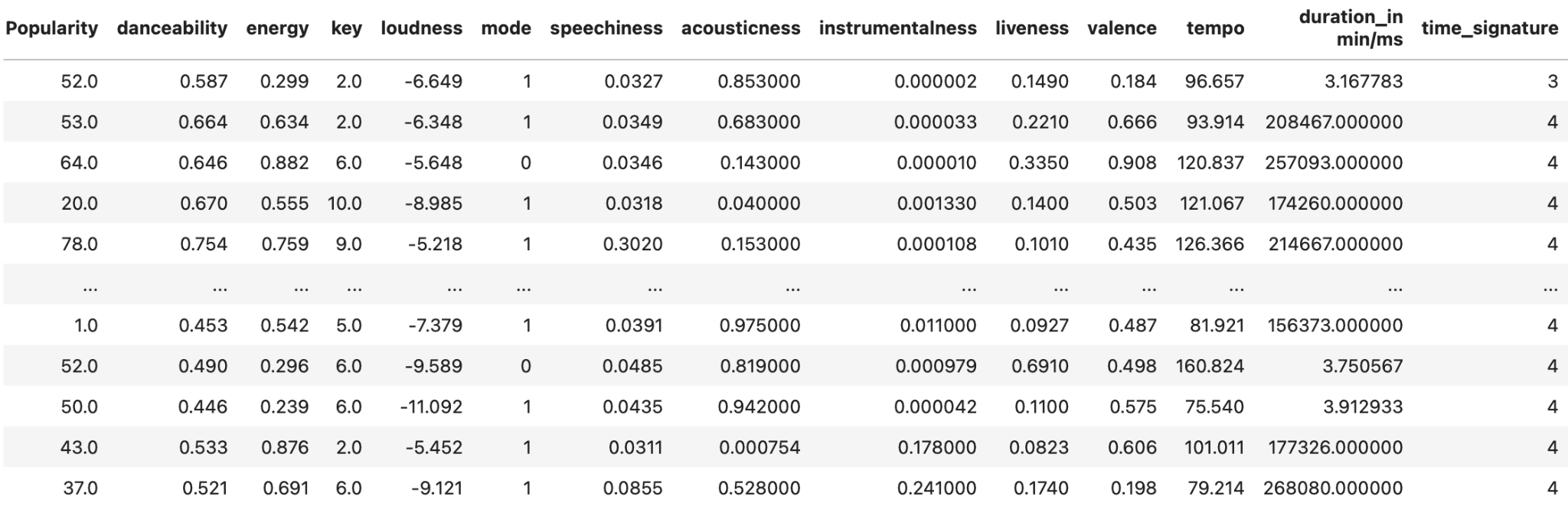
Final Project

Music Genre Classification

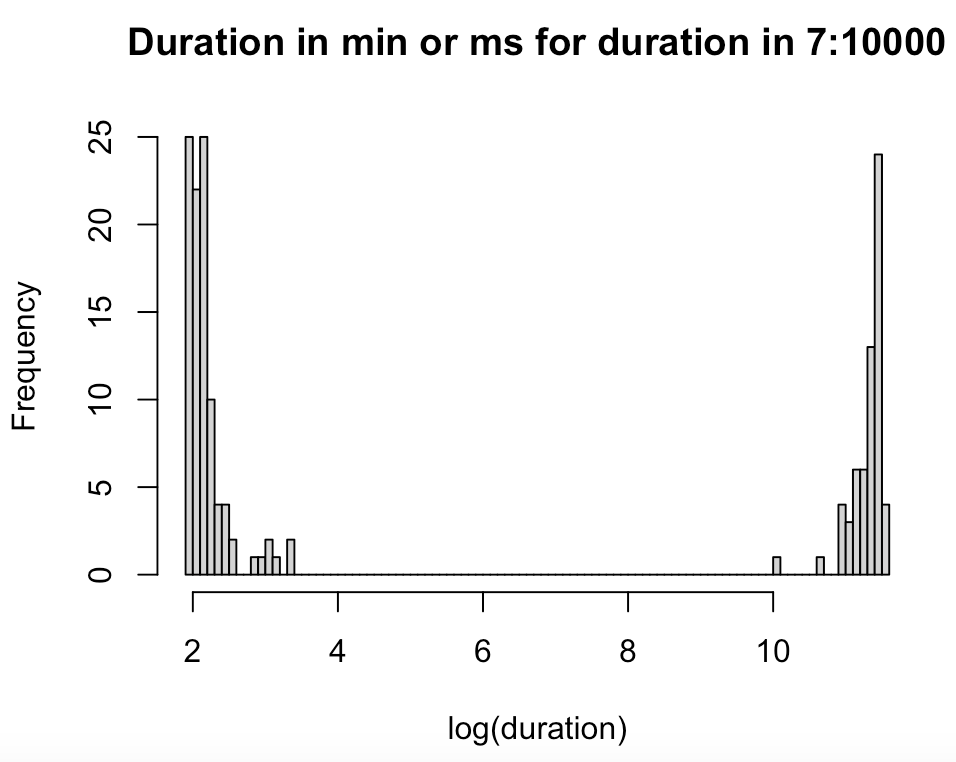
The goal of this project is to classify the genre of songs. The dataset can be accessed by visiting the kaggle website in the works cited and downloading the attached train.csv file (the test.csv file has no labels).

This model may be useful for apps like spotify or apple music as part of their algorithm that decides which songs or playlists to recommend to the user. For example, if a spotify user listens to a lot of pop music but is searching for country songs, it may be beneficial to suggest country songs that the model classifies as pop songs. This way, the country songs that are suggested are likely to sound like the pop songs that the app knows the user enjoys.

The features contain information about each song, including popularity, danceability, energy, key, loudness, mode, speechiness, acousticness, instrumentalness, liveliness, valence, tempo, duration, and time signature. After removing rows with NA values, there are 11,113 rows in the features, with each row corresponding to a song. A glimpse of the features is shown below.

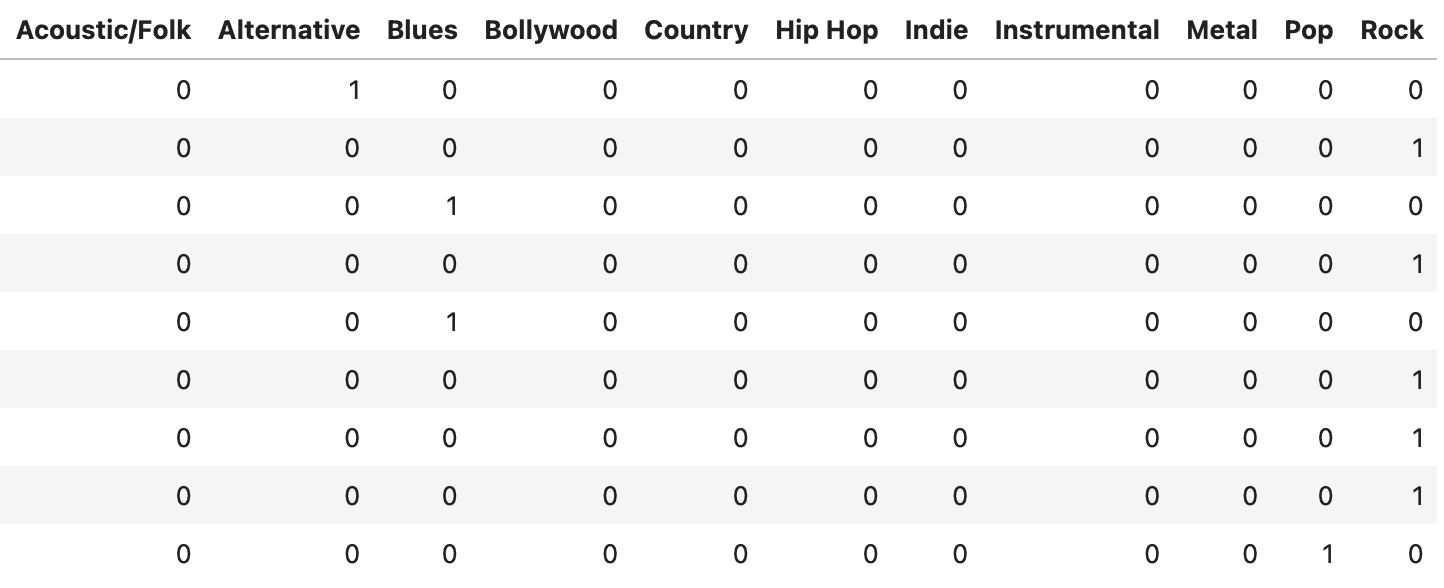


The columns “key” and “time\_signature” are categorical but are labeled as integers. Those features are one-hot-encoded before neural network training. The column “duration\_in min/ms” is the duration of a song in either minutes or milliseconds. Fortunately, it is easy to differentiate between the units, as seen in the graph below of the frequencies of songs for durations between 7 and 10000 .



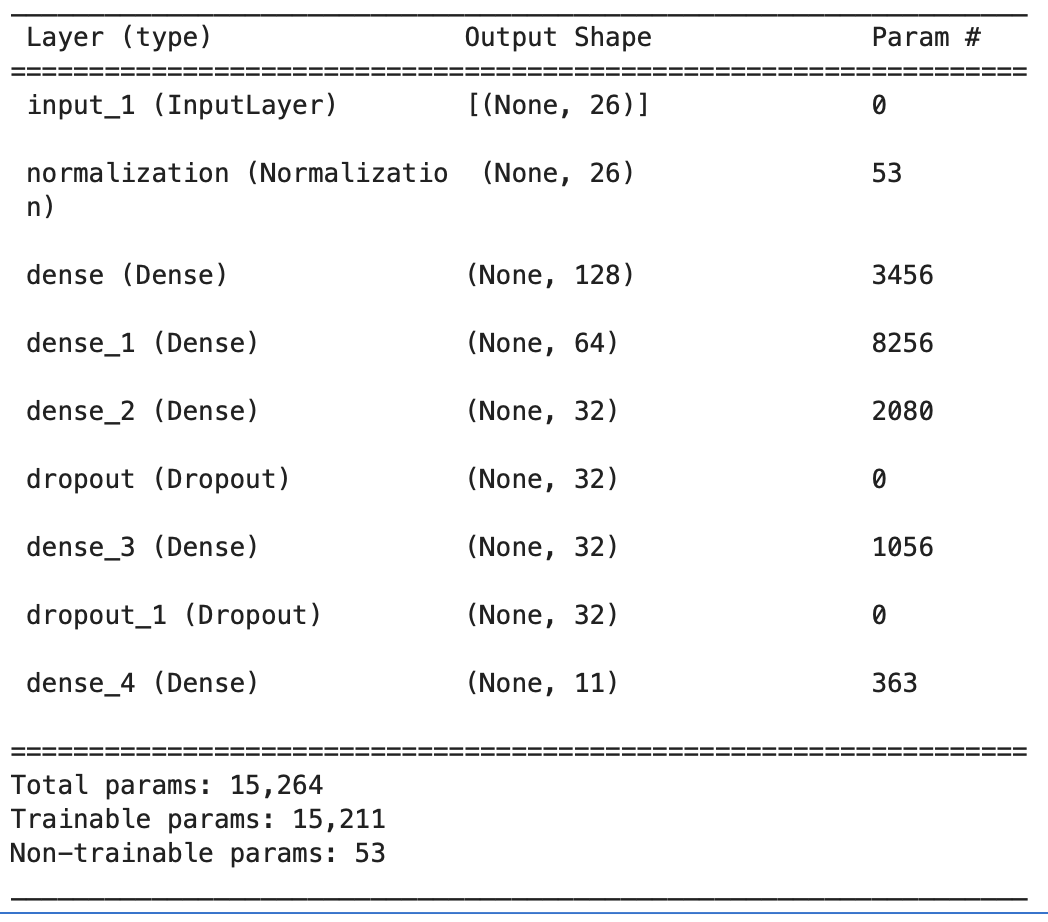
The large gap in the data surely separated the songs measured in minutes from the songs measured in milliseconds. All durations less than 403 (e^6) were multiplied by 60,000 so that all observations had the same unit (milliseconds).

The labels are 11 genres that are shown as one-hot encoded in the table below.

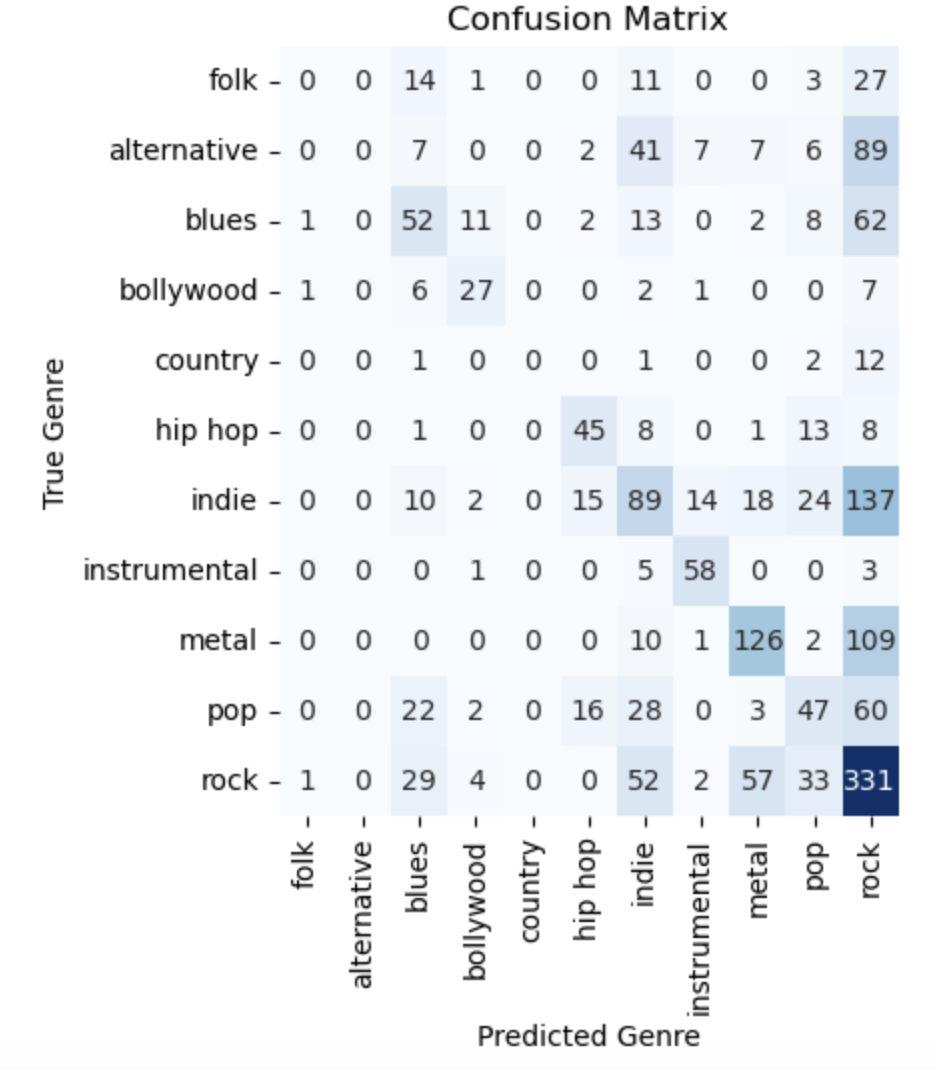


**Model Fitting**

A look at the structure of the neural network is shown below.

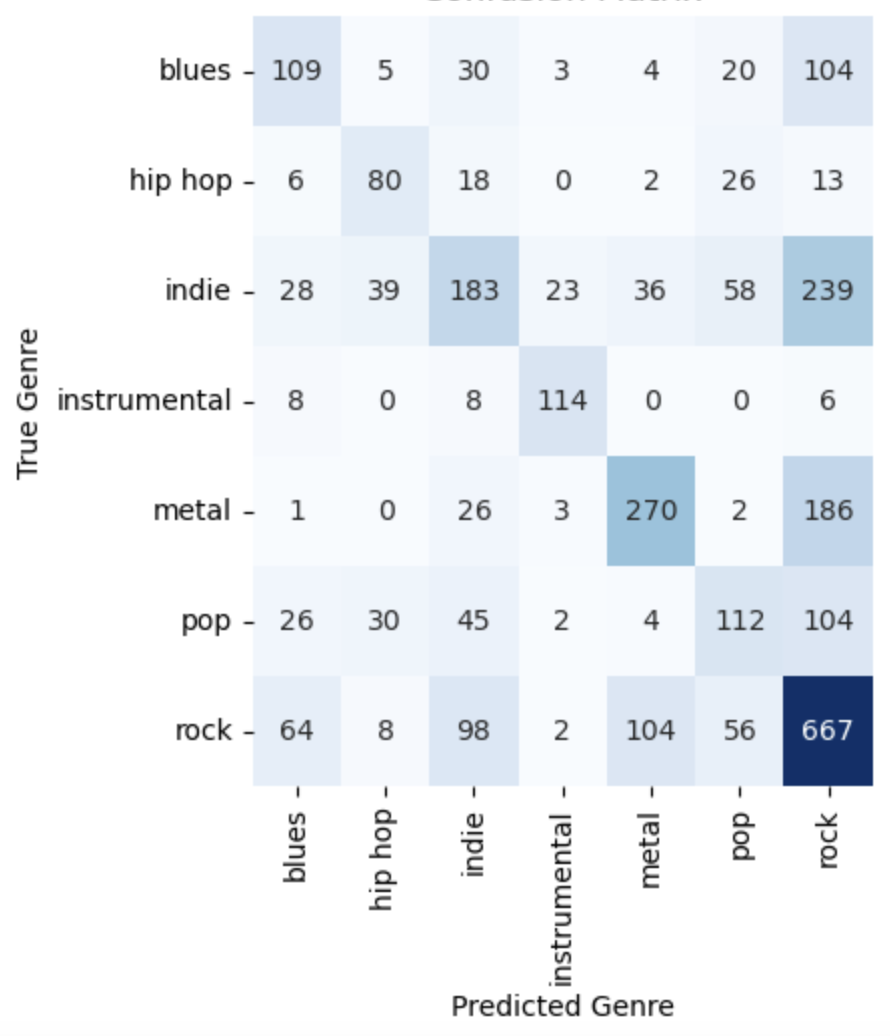


After some tuning, a network was trained for 200 epochs using the ADAM optimizer with a learning rate of 0.00001, a batch size of 1, and a validation split of 20% of the data. An estimated test accuracy of 0.4275 was obtained. While this test accuracy is considerably better than a random guess (0.09), it is not ideal and even lower than the test accuracy obtained by the person who uploaded the dataset to Kaggle (0.53). Below is a confusion matrix of the model’s predictions on the test set.



We can see that the network predicted Rock many times and often incorrectly. This makes sense because Rock is the most popular genre in the dataset. No song was predicted to be Country or Alternative.

A higher accuracy can likely be obtained by reducing the number of genres. A network was trained for a dataset of songs without the folk, alternative, bollywood, or country genres. These genres were chosen to be excluded because there are not many instances of them in the dataset. The estimated test accuracy of the new model is 0.5165. The confusion matrix on the test set is shown below.



**Works Cited**

“Confusion\_matrix.” *Scikit*, scikit-learn.org/stable/modules/generated/sklearn.metrics.confusion\_matrix.html. Accessed 14 May 2025.

“Statistical Data Visualization#.” *Seaborn*, seaborn.pydata.org/. Accessed 14 May 2025.

Malgi, Purushottam. “Music Genre Classification.” *Kaggle*, 7 Aug. 2021, www.kaggle.com/datasets/purumalgi/music-genre-classification/code.