TJ Tracy
FP
DAT 350
4/20/24

Music Genre Classification

I. Introduction

The goal of the project is to implement a machine learning approach to classify the genre of music given data on the audio files. The approach uses Decision Tree, Logistic Regression, Support Vector Machine, Neural Network, Random Forest, and k-Nearest Neighbors (which will be referred to as k-NN) to tackle the multi-class problem of music genre classification. The data for the classification is a collection of features that describe the audio track. There are a total of 1000 audio tracks and each 30 seconds in length. Each genre is represented by 100 different audio tracks and is labeled with the correct genre of music for that audio file. Therefore, the problem is well-suited for supervised learning approaches.

II. Data, Preprocessing, & Design

There are two sets of data available. The first includes 1000, 30-second-long audio files, where each genre is represented by 100 different tracks. The original 30-second-long clips are divided into audio files 3-seconds in length. That process increases the number of observations by approximately ten-fold. As a result, the dataset that the algorithms are trained on contain 9990 rows. There are total of 60 features that quantify the structure of the song and are represented by the variance and the mean. The data contains no missing data and all features that represent the mean are relatively normally distributed. While the features measuring variance are right skewed. The original label of the music is label encoded since there is no implied order to the music label encoding is appropriate and the use of One Hot Encoding is unnecessary.

The dataset is first divided into a training and testing set, where the training set is 95% of the original dataset. Then the training is set is divided into a training and validation set where the validation set is 20% of the original training set. Reminder that the dataset has many columns at

60, therefore wrapper feature selection is employed to select the best combination of 10 variables to classify the music. After the data has been processed it is ready to be used in hyperparameter optimization. Hyperparameter optimization is an iterative process that trains multiple models with different parameters. After evaluating all approaches, the parameters that returned the best results are selected. The models in the hyperparameter tuning stage are cross validated five times and after the cross-validation the best parameters are returned.

The best parameters for each model are then trained and evaluated using 5-fold cross-validation. Within the cross-validation process the features are scaled using MinMax Scaling which rescales the data to have a range of 0 to 1. Notably, there are no negative values, so the range is appropriate for the data. After cross-validation, the best performing models are:

A. Neural Network
B. Random Forest
C. k-NN

To this point the models have been trained on a training and validation set that is separate from the final testing set. This is done to ensure the models have not been exposed to the final testing set and thus returning a true evaluation of how the models perform on data it has not been exposed to before. The performance of the models is evaluated using precision, recall, f1 score, and accuracy.

III. Results

Each model is retrained and tested on the testing set that has been separate from the training and validation process. The model that performed the best is k-Nearest Neighbors (See Appendix II for results from the Neural Network and the Random Forest). The k-NN performed well, achieving a macro average and an accuracy of 92% (See Table 1 Below). On average, the model can identify 92% of the instances of the target class and when the model predicts the genre of music it is correct 92% of the time. The model performed best when predicting blues music and performed the worst on classical music. Compared to previous k-NN results in the literature the model outperforms the most comparable results by approximately 34% in terms of accuracy when predicting across all genres excluding classical and jazz.

Table 1: k-NN Results

Genre	Precision	Recall	F1 Score	Support
Rock	0.94	0.96	0.95	50
Pop	0.96	0.94	0.95	50
Reggae	0.92	0.88	0.90	50
Jazz	0.85	0.94	0.90	50
Country	0.94	0.92	0.93	50
Hip-Hop	0.90	0.90	0.90	50
Blues	0.98	0.98	0.98	50
Metal	0.96	0.94	0.91	50
Disco	0.89	0.94	0.91	50
Classical	0.91	0.84	0.88	50
Accuracy			0.92	500
Macro Average	0.92	0.92	0.92	500
Weighted Average	0.92	0.92	0.92	500

Table 2: Previous Results (Tzanetakis & Cook 2002)

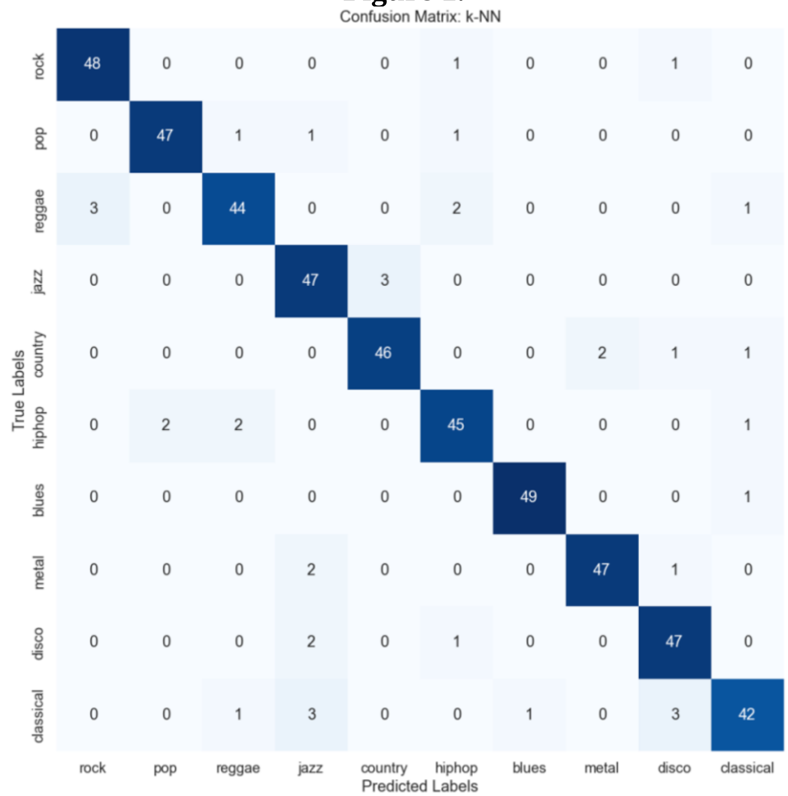
CLASSIFICATION ACCURACY MEAN AND STANDARD DEVIATION

	Genres(10)	Classical(4)	Jazz(6)
Random	10	25	16
RT GS	44 ± 2	61 ± 3	53 ± 4
GS	59 ± 4	77 ± 6	61 ± 8
GMM(2)	60 ± 4	81 ± 5	66 ± 7
GMM(3)	61 ± 4	88 ± 4	68 ± 7
GMM(4)	61 ± 4	88 ± 5	62 ± 6
GMM(5)	61 ± 4	88 ± 5	59 ± 6
KNN(1)	59 ± 4	77 ± 7	57 ± 6
KNN(3)	60 ± 4	78 ± 6	58 ± 7
KNN(5)	56 ± 3	70 ± 6	56 ± 6

TJ Tracy
FP
DAT 350
4/20/24

IV. Appendix 1 (Correlation Matrices)

Figure 1.



[1]

Figure 2.

GENRE CONFUSION MATRIX

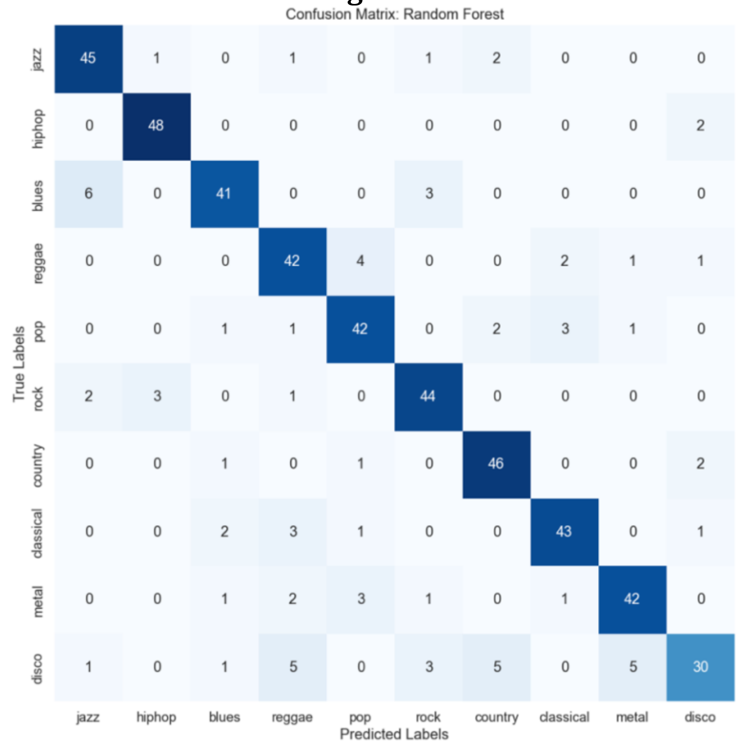
	cl	co	di	hi	ja	ro	bl	re	po	me
cl	**69**	0	0	0	1	0	0	0	0	0
co	0	**53**	2	0	5	8	6	4	2	0
di	0	8	**52**	11	0	13	14	5	9	6
hi	0	3	18	**64**	1	6	3	26	7	6
ja	26	4	0	0	**75**	8	7	1	2	1
ro	5	13	4	1	9	**40**	14	1	7	33
bl	0	7	0	1	3	4	**43**	1	0	0
re	0	9	10	18	2	12	11	**59**	7	1
po	0	2	14	5	3	5	0	3	**66**	0
me	0	1	0	1	0	4	2	0	0	**53**

[2]

[1] Results from k-NN approach.
[2] Results from Tzanetakis & Cook (2002).

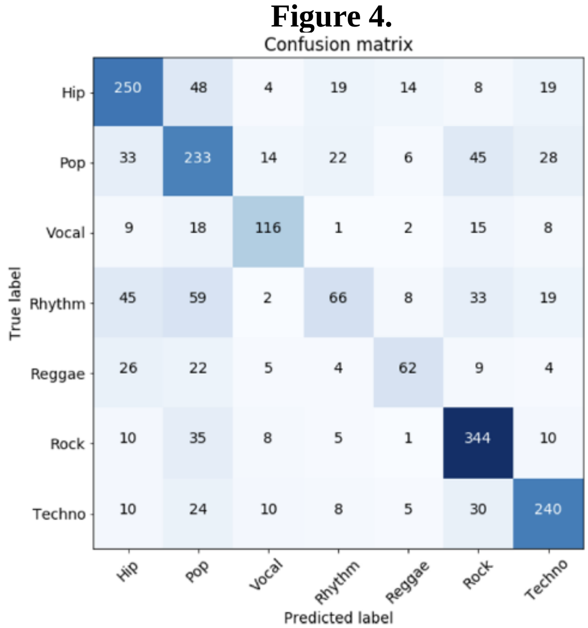
TJ Tracy
FP
DAT 350
4/20/24

Figure 3.



[3]

Figure 4.



[4]

---

[3] Results from Random Forest approach.
[4] Results from Ensemble approach Bahuleyan (2018).

TJ Tracy
FP
DAT 350
4/20/24

V. Appendix 2 (Further Results)

Table 2: Neural Network

	Precision	Recall	F1 Score	Support
Rock	0.90	0.88	0.89	50
Pop	0.94	0.96	0.95	50
Reggae	0.78	0.76	0.77	50
Jazz	0.91	0.82	0.86	50
Country	0.90	0.90	0.90	50
Hip-Hop	0.82	0.90	0.86	50
Blues	0.92	0.92	0.92	50
Metal	0.94	0.9	0.92	50
Disco	0.87	0.9	0.88	50
Classical	0.75	0.76	0.75	50
Accuracy			0.87	500
Macro Average	0.87	0.87	0.87	500
Weighted Average	0.87	0.87	0.87	500

Table 3: Random Forest

	Precision	Recall	F1 Score	Support
Rock	0.85	0.88	0.86	50
Pop	0.82	0.84	0.83	50
Reggae	0.76	0.84	0.80	50
Jazz	0.83	0.9	0.87	50
Country	0.84	0.92	0.88	50
Hip-Hop	0.92	0.96	0.94	50
Blues	0.87	0.82	0.85	50
Metal	0.86	0.84	0.85	50
Disco	0.83	0.60	0.70	50
Classical	0.88	0.86	0.87	50
Accuracy			0.85	500
Macro Average	0.85	0.85	0.84	500
Weighted Average	0.85	0.85	0.84	500

TJ Tracy
FP
DAT 350
4/20/24

References

Bahuleyan, H. (2018). Music Genre Classification using Machine Learning Techniques. *ArXiv, abs/1804.01149*.

Haggblade, M., Hong, Y., & Kao, K. (2024). Music Genre Classification. *International Research Journal of Modernization in Engineering Technology and Science*.

Tzanetakis, G., & Cook, P.R. (2002). Musical genre classification of audio signals. *IEEE Trans. Speech Audio Process., 10*, 293-302.