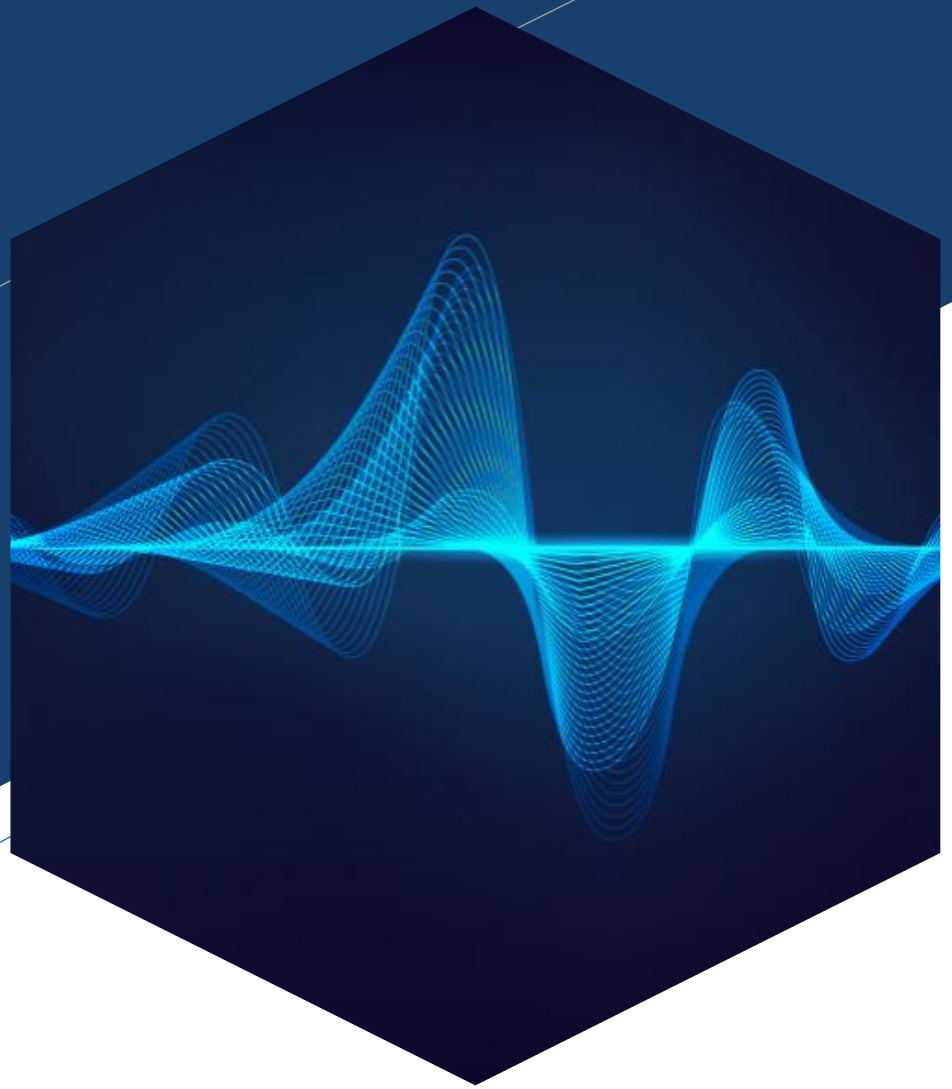


Music Genre Classification

Capstone Project 2

Ishmail Grady

September 2019

Project Proposal

Music Genre Classification

- Music classification is a key problem in machine leaning that can enhance technology that interacts with music to gain insight or perform various tasks such as music generation, track identification and playlist construction
- The goal of this project is to create a machine learning model that is designed to take audio features of a MP3 file and predict the musical genre of the track.
- In doing so, this project will explore the advantages and disadvantages of both approaches to determine the best solution

Project Approach

Music Genre Classification

- To provide a basis for determining if deep learning models provide a significant advantage over non-deep learning models , three models were developed:
  - Deep Learning Classifier – this model will be trained on mel-spectrograms of audio files
  - Baseline Ensemble Classifier – this model will be trained on pre-computed audio features
  - Stacked Ensemble Classifier – this model will be trained on pre-computed audio features and the predicted probabilities of the Deep Learning Classifier
- Comparing the performance of the Deep Learning Classifier to the baseline ensemble classifier will determine if the deep learning model provides significant advantages over non-deep learning models.
- The stacked ensemble classifier will determine if using the predictions from the deep learning model can significantly enhance the performance of the baseline ensemble classifier.

Data Wrangling

Data Source

Music Genre Classification

- The main data source for this project will come from a compilation of high-quality audio files that are free and legal to use through the Free Music Archive (FMA) which is designed for musical analysis
- This dataset contains 8,000 30 second track samples from 8 different genres. The FMA has untrimmed tracks for over 100,000 songs, however, due to memory and computing power constraints, this project will only use a small subset
- The dataset contains metadata on each track and pre-computed audio features which are publicly available for download.

Data Wrangling & Cleaning

Music Genre Classification

- This project uses two datasets for the same audio files:
  - Mel-spectrogram arrays - the 30 sec raw MP3 audio files were converted* to mel-spectrograms and were stored as numerical arrays The mel-spectrogram arrays were used to train the deep learning models.
  - Precomputed audio features – this dataset was constructed from 3 main tables: Tracks (track metadata and response labels), features (pre-computed audio features), and genre (genre labels and hierarchy) The precomputed audio features were used to train the non-deep learning models.
- Metadata such as artist name, number of listens, and year the track was made were removed from the dataset.

*6 of 8,000 tracks were not able to be converted and were discarded from the dataset

Models

Deep Learning Classifiers

Music Genre Classification

- A convolutional neural network (CNN) and a convolutional recurrent neural networks (CRNN) were considered
- CNNs are useful for dynamically learning spatial features of an image, and in this application, features of a mel-spectrogram.
- CRNNs combine the advantages of CNNs with the advantages of recurrent neural networks which have the ability to improve predictions by modeling a temporal sequence of extracted features, thus future inputs use information from previous inputs
- The test set was used to compare the performance of each model to determine the best deep learning model for this project

*6 of 8,000 tracks were not able to be converted and were discarded from the dataset

Ensemble Classifiers

Music Genre Classification

- The Baseline and Stacked Ensemble Classifiers share the same structure
  - The ensemble voting classifiers are composed of a logistic regression classifier, a gradient boosting random forest classifier, and a support-vector machine classifier.
  - The voting method of both models is ‘soft’ meaning that it predicts class labels based on the argmax of the sums of the predicted probabilities
- The Stacked Ensemble Classifier is trained on both the precomputed audio features and the predictions from the final Deep Learning Classifier model

*6 of 8,000 tracks were not able to be converted and were discarded from the dataset

Results

Deep Learning Classifiers

Music Genre Classification

Dataset Split	CNN	Parallel CRNN
Training	.43	.53
Validation	.34	.42
Testing	.25	.36

- The Parallel CRNN had the best performance on the test data with a macro-F1 score of .36.
- This is a significant improvement on the CNN model which achieved a macro-F1 score of .25.
- The Parallel CRNN was used as the final deep learning model and the predictions form this model will be used as features for stacked ensemble classifier.

Ensemble Classifiers

Music Genre Classification

Dataset Split	Baseline Ensemble Classifier	Stacked Ensemble Classifier
Logistic Regression	.44	.46
Gradient Boosting	.47	.49
SVC	.49	.50
Overall	**.49**	**.51**

- The stacked ensemble voting classifier was has test a macro-F1 score of .51 which is .02 higher than the baseline ensemble voting classifier.
- The use of CRNN predictions as features was able to significantly improve prediction performance.
- The SVC outperformed the gradient boosting and logistic regression models for both classifiers. The SVC actually outperformed the full baseline ensemble classifier. The full stacked classifier only improved upon the accuracy of the SVC model by 0.05.

Final Model Selection

Music Genre Classification

- The highest performing model for this project with a goal of creating a predictive model for music genre classification was the Stacked Ensemble Classifier with a macro-F1 score of .51.
  - This model would be most useful for predicting Hip Hop, Electronic and Rock genres which are 3 of the most popular genres in modern society.
- In practice, it is recommended to use the SVC model of the baseline ensemble classifier when speed and computing power are important considerations.
  - Even though the stacked ensemble voting classifier is more accurate, this gain in accuracy may not be worth the extra computing resources that it takes to process data and train the model of the deep learning classifier so that its predictions can be used as features

Conclusion

Music Genre Classification

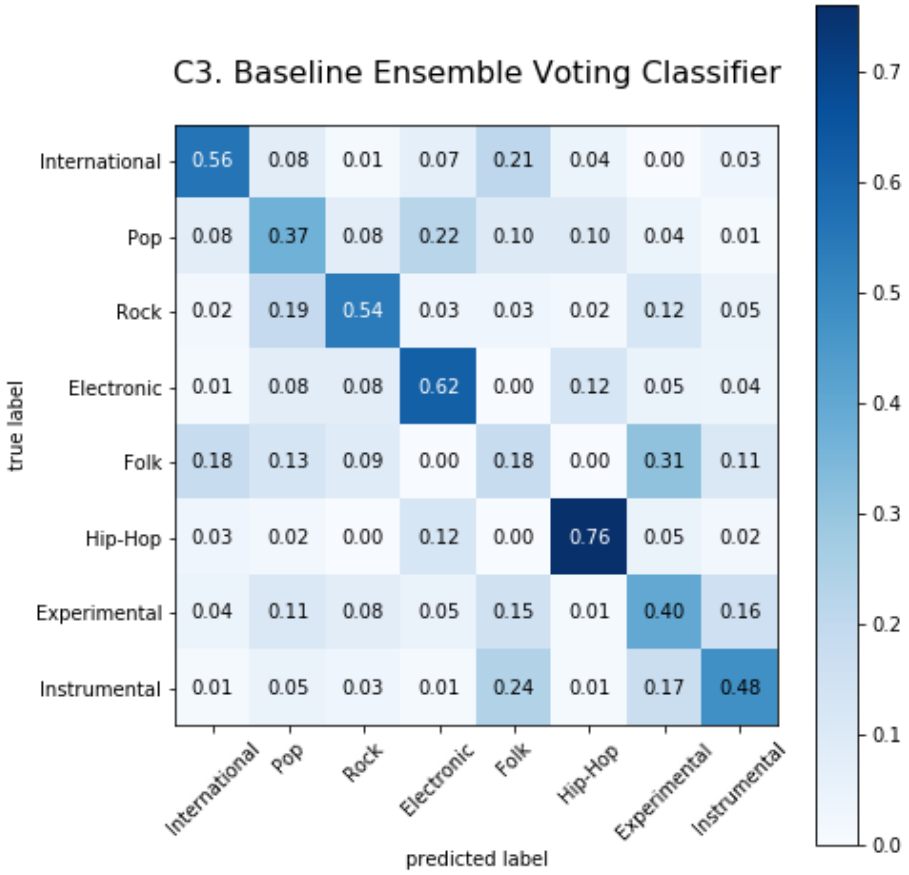
- It can be concluded that the use of the predictions from the deep learning classifier can improve the performance of traditional ML methods as shown by the performance of the stacked ensemble classifier.
- In practice, it is recommended to use the SVC model of the baseline ensemble classifier when speed and computing power are important considerations.
- It is important to note that these conclusions can only be made under the specific constraints and parameters of this project. It is expected that the accuracy of the deep learning models will increase when trained on more data.

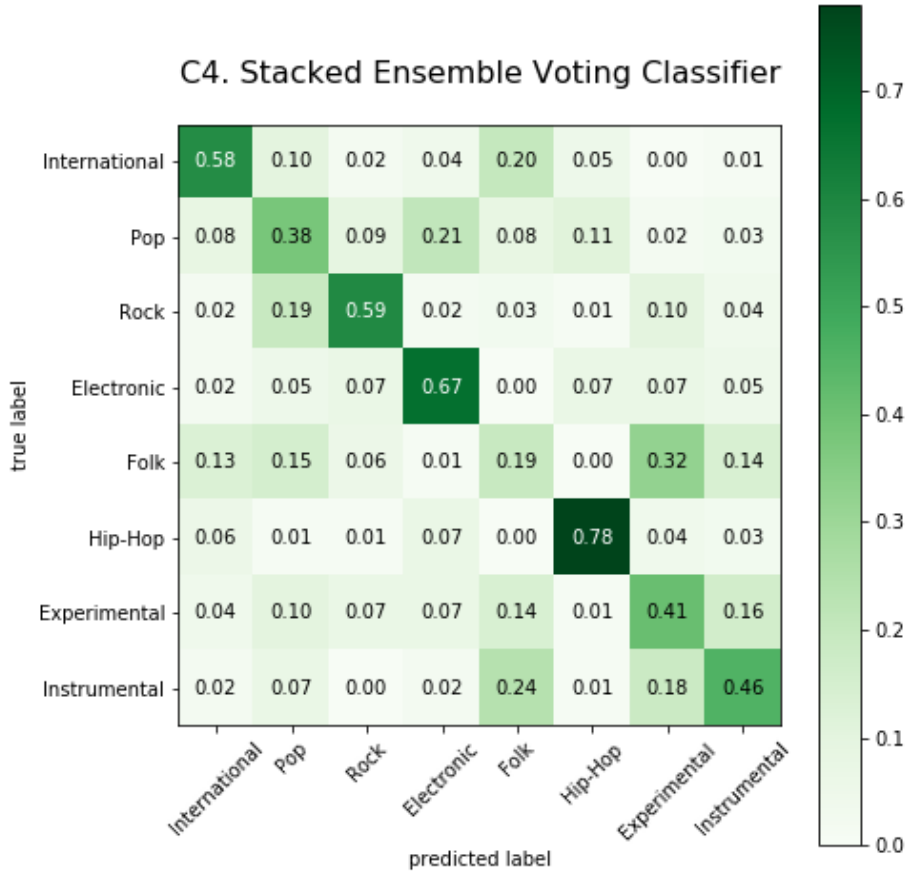
Appendix

Ensemble Classifier Results

Music Genre Classification

Normalized Confusion Matrices





Ensemble Classifier Results

Music Genre Classification

F1 Scores by Genre

Models	Baseline Ensemble Classifier	Stacked Ensemble Classifier
International	.577	.594
Pop	.373	.354
Rock	.584	.615
Electronic	.588	.632
Folk	.199	.206
Hip Hop	.725	.769
Experimental	.370	.391
Instrumental	.492	.476