Music Genre Classification

Louis Gomez LGOMEZ@STEVENS.EDU *Department of Computer Science*
Stevens Institute of Technology
Hoboken, New Jersey

1. Introduction

In recent years, music is increasingly a confluence of different sounds derived from a variety of genre which makes categorizing them for things like award shows and musical charts. In this project, we aim to computationally devise what four difference genres of music are and develop machine learning models to classify them. To do this, we compare two methods: a simple K Nearest Neighbour and a Random Forest.

2. Experiment Design

2.1. Data Collection

For the classification task of genre classification, we used music data from Spotify, a music streaming platform that has about 30 million songs. Our classes of interests are Rap, Country, Electronic Dance Music (EDM) and RnB. For selecting the songs in these categories we focused on music playlists that already had genre labels. We extracted music from nine popular Spotify curated playlists and in total we had 509 songs. We had 102 Country, 158 EDM, 129 Rap and 120 RnB songs. For data collection, we use Spotify's music API to query their servers to get songs in each playlist Spotify (2019).

2.2. Feature Extraction

To extract features for each song, We utilize Spotify's audio feature service that returns audio level features curated by Spotify. In total, there are 14 features retrieved. but we utilize only 12 of these features. These features include: key, mode,acousticness, energy,danceability, instrumentalness, liveness, loudness, speechiness, valence,tempo, and time signature. The 2 features not utilized are the time duration and song URL. These features logically provide no valuable information as to whether a song belongs to a specific genre.

2.3. Modelling

For this project, I utilized a machine learning library called Scikit Learn Pedregosa et al. (2011).We sought to compare the accuracy between a simple learner, KNN and a complex learner, Random Forest. We choose KNN because songs within the same genre tend to be closer together and this could be particularly useful since KNN relies on distance. Conversely, we choose Random Forest due to it robust nature and ability to form complex relationships between features. For both of these cases, we standardize each of our features to have a mean of zero and variance of one. For our categorical labels, we label encode them from [Country, EDM, RB, Rap] to $[0, 1, 2, 3]$. We split our entire data set into a train and test split, with the train set containing 70 percent and test set containing the remaining 30 percent.

K Nearest Neighbour KNN is a machine learning algorithm that uses the K nearest neighbours (distance) to a data point to assign labels to each data point. The main hyper parameter in KNN is the K parameter. To obtain the optimal value for K, we performed a 10-fold cross validation on a range of K values from 1 - 30 over the entire training data. The estimator with the highest cross validation accuracy (0.657) was chosen and it used K = 26. To perform cross validation, we utilized the Cross Validation and Grid Search libraries. To train our KNN classifier, we utilized Scikit Learn KNeighbourClassifier Pedregosa et al. (2011)

Random Forest As with KNN, to tune the hyper parameters of the Random Forest, we utilize a random grid search and cross validation to tune hyper-parameters. We used a Randomized grid search library due to the large computation time required to tune every single hyper parameter. We settled on a 100 trees, a maximum depth of 5, minimum sample split of 5 and a minimum sample in each leaf as 2. We trained and tested using the Random Tree Classifier Pedregosa et al. (2011).

3. Experimental Results

To evaluate the classification accuracy of our classifier, we calculated the precision, recall and f1 scores for each class on the test set. Contained in the test set are 25 Country labels, 46 EDM labels, 45 Rap labels and 37 RnB labels.

KNN Using KNN on the test set and as shown in Table 1 , we achieved an overall average accuracy of 0.65. Further examining the individual precision, recall and f1 scores, the highest f1 score was seen in EDM class with 0.74. The highest recall was seen in the Country class with 0.92 while the highest precision was seen in the Rap class with 0.89. Additionally, the RnB Class had a low recall and precision of 0.54. This means that of all the 37 samples in the test set labelled as RnB, it was accurate 54% of the time. On precision, the Country class scored lowest with 0.49 which means that of all the test samples labelled as country, only 49% of them were actually of the country class.

Random Forest Using Random Forest, we improve performance across the four classes. We achieve higher precision, recall and f1 score in most of the genre classes as seen in Table 1 but we have marginal reductions in some. Notably our classification performance on the RnB class is increased from 0.54 using KNN to 0.61 using random forest. Additionally, the precision on the country class improved from 0.49 using KNN to 0.65. Our overall average accuracy was 0.73. For feature importance, the top five important features for the random forest were: speechiness, energy, instrumentalness, danceability, and tempo.

Result Analysis To get a better understanding of where classification occurs, we plot confusion matrices for both models as seen in figure 2. From figure 2.1 we can visualize the results of the KNN model. For example, we can see that in the Rap class, 25 songs in the rap class are predicted correctly but 20 songs are misclassified which leads to a recall of 0.56. Also the country class has a high recall because it predicted 23 out of the 25 songs in

	KNN			Random Forest			
	Precision	*Recall*	*F1Score*	*Precision*	*Recall*	*F1 Score*	*No. in Test set*
Country	0.49	0.92	0.64	0.65	0.88	**0.75**	25
EDM	0.78	0.70	0.74	0.76	0.74	**0.75**	46
RnB	0.54	0.54	0.54	0.63	0.59	**0.61**	37
Rap	0.89	0.56	0.68	0.87	0.76	**0.81**	45

Figure 1: Summarized Results from the KNN and Random Forest Models

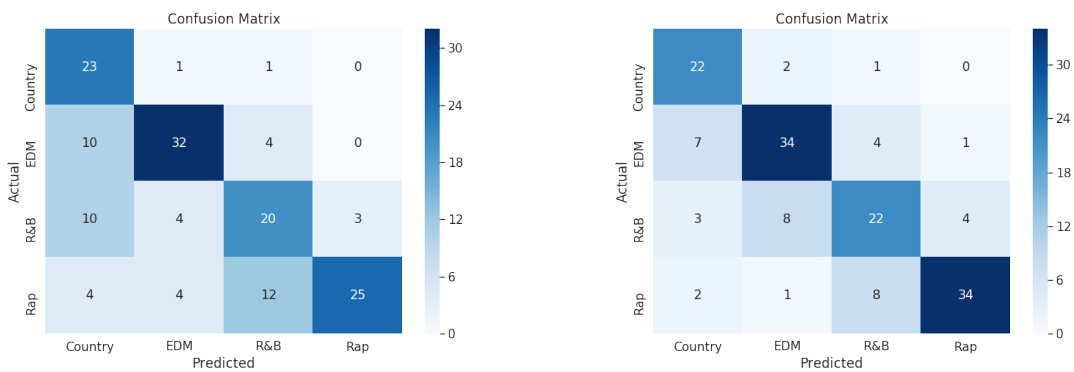

Figure 2: Left: A heat-map visualizer for the confusion Matrix of KNN and RF Models. The left Plot if the KNN while the right plot is the Random Forest

its class to be the right label. We can also see similar relationships and understanding in the Random Forest confusion matrix as well.

4. Conclusion

In summary, we have shown that is feasible to predict the genre of music given a common set of features. Using our simple model KNN we achieved an overall accuracy of 0.65. Using a more advanced method such as Random forest improves this accuracy to 0.73. For future work, it be advantageous to investigate the misclassified songs and see why they were said to be of specific genres.

References

F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

Spotify. Spotify web api, 2019. URL https://developer.spotify.com/documentation/web-api/reference-beta/.