Recognition and Classification of Emotions in Music

Aditya Nigam (11040) and Revant Teotia (13564)

Group - 2

Mentor: Dr. Piyush Rai

Department of Computer Science and Engineering, IIT Kanpur

Introduction:

- Music is an integral part of human life. Often, music is associated with important moments of our life, brings to us memories and evokes emotions.
- Due to frantic increase in the amount of music available these days, classification and recognition of emotions conveyed by music has become indispensable in todays world.
- No classification algorithm has been able to generate great accuracy on these dataset.
- We have attempted to perform a comparative study of the different classifiers and their ability to predict different genres with varied accuracy.

Dataset:

- ✓ Dataset has been prepared from All Music.com dataset, which consists of 903 audio clips of 30 seconds each.
- ✓ Dataset has been taken from http://mir.dei.uc.pt/resources/MIREX-like_mood.zip made available by Renato Panda, Bruno Rocha and Rui Pedro Paiva.
- ✓ The dataset is in accordance with the MIREX dataset which is the base of comparison generally accepted by the music emotion recognition community.

Genre Clusters:

- **Cluster 1** • passionate, rousing, confident, boisterous, rowdy
- **Cluster 2** • rollicking, cheerful, fun, sweet, amiable/good natured
- **Cluster 3** • literate, poignant, wistful, bittersweet, autumnal, brooding
- **Cluster 4** • humorous, silly, campy, quirky, whimsical, witty, wry
- **Cluster 5** • aggressive, fiery, tense/anxious, intense, volatile, visceral

Features extracted:

- **RMS:** Root mean Square approximates the loudness of the sound. It is calculated by taking RMS of the amplitudes of the spectrum of sound.
- **Mel-Frequency Cepstral Coefficients:** MFCC represents a set of short term power spectrum characteristics of the sound. It models the characteristics of human voice. It can be derived as follows:
- ✓ Take the Fourier transform of (a windowed excerpt of) a signal.
- ✓ Map the powers of the spectrum obtained above onto the mel scale, using triangular
- ✓ overlapping windows.
- ✓ Take the logs of the powers at each of the mel frequencies.
- ✓ Take the discrete cosine transform of the list of mel log powers, as if it were a signal.
- ✓ The MFCCs are the amplitudes of the resulting spectrum.
- **Spectral Flux:** Spectral flux is a measure of how quickly the power spectrum of a signal is changing, calculated by comparing the power spectrum for one frame against the power spectrum from the previous frame.
- ✓ It is usually calculated as the Euclidean distance between the two normalised spectra. The spectral flux is not dependent upon overall power and phase.
- **Spectral Mean:** This is average of all the frequencies in a spectrum. Unlike centroid it is not calculated by assigning weights to the frequencies
- **Spectral Centroid**: The spectral centroid is a measure used in digital signal processing to characterise a spectrum. It indicates where the 'center of mass' of the spectrum is. It is calculated as:

$$Centroid = \frac{\sum_{n=0}^{N-1} f(n)x(n)}{\sum_{n=0}^{N-1} x(n)}$$

- ✓ where x(n) represents the weighted frequency value, or magnitude, of bin number n, and f(n) represents the center frequency of that bin.
- **Zero Crossing Rate:** Zero Crossing Rate is the zero crossing count (ZCC) in which it counts the number of times the sign changes from positive to negative or vice versa per second in a signal.
- **Rolloff:** Spectral Rollo point is the frequency below which the 85 % of the magnitude of the distribution is concentrated. The equation to calculate rolloff is:

$$\sum_{n=1}^{M} |x[n]| = 0.85 \sum_{n=1}^{N/2} |x[n]|$$

- **Skewness:** From skewness we can know, how much the shape of the spectrum below the Spectral Centroid is different from the shape above. For a white noise, the skewness is zero.
- **Variance:** It is variance of frequencies from the spectral centroid. Variance gives us a measure for how much the frequencies deviate from the Spectral Centroid in a spectrum.
- **Flatness:** Spectral Flatness measures the atness of a spectrum. It is also used to distinguish between noise-like and tone-like sounds and calculated using the equation:

$$Flatness = \frac{\sqrt[N]{\prod_{n=0}^{N-1} x(n)}}{\frac{\sum_{n=0}^{N-1} x(n)}{N}} = \frac{exp(\frac{1}{N}\sum_{n=0}^{N-1} ln x(n))}{\frac{1}{N}\sum_{n=0}^{N-1} x(n)}$$

- **Spectral Crest Factor:** Spectral Crest factor is the ratio of the maximum spectrum power and the mean spectrum power of a sub-band. It measures the peakness of a spectrum. It is used to distinguish between noise-like and tone-like sounds.

Results:

- ➢ Confusion matrix before and after merging of clusters for SVM Classier:

		Predicted Label				
		C1	C2	C3	C4	C5
True Label	C1	28.12	6.25	21.87	18.75	25
	C2	5.26	21.05	23.68	34.21	15.78
	C3	0	0	78.37	16.21	5.4
	C4	12.19	9.75	26.82	36.58	14.63
	C5	15.62	6.25	9.38	15.62	53.12

		Predicted Label		
		C1 + C5	C2 + C4	C3
True	C1 + C5	60.93	23.43	15.62
	C2 + C4	24.05	50.63	25.32
	C3	5.40	16.22	78.38

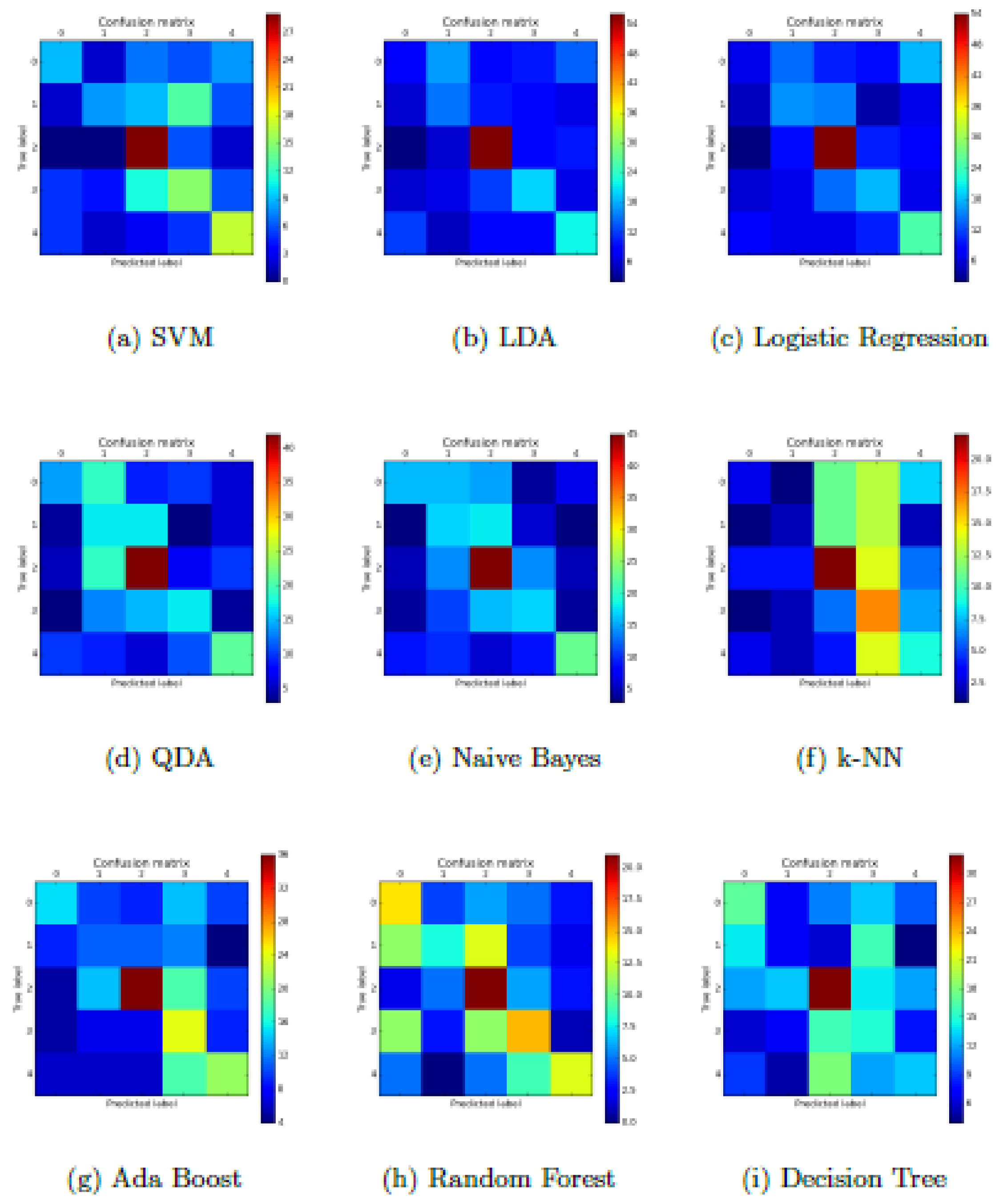
- ➢ Confusion matrix before and after merging of clusters for Ensemble Classifier:

		Predicted Label				
		C1	C2	C3	C4	C5
True Label	C1	20.68	27.58	17.24	12.06	22.41
	C2	12.7	29.78	23.4	17.02	17.02
	C3	2.4	10.84	67.46	8.43	10.84
	C4	3.8	11.53	30.76	38.46	15.38
	C5	19.14	14.89	14.89	25.53	25.53

		Predicted Label		
		C1 + C5	C2 + C4	C3
True	C1 + C5	43.81	40.95	16.19
	C2 + C4	25	48	47
	C3	13.25	19.28	67.47

- ➢ Statistics for different classifiers:

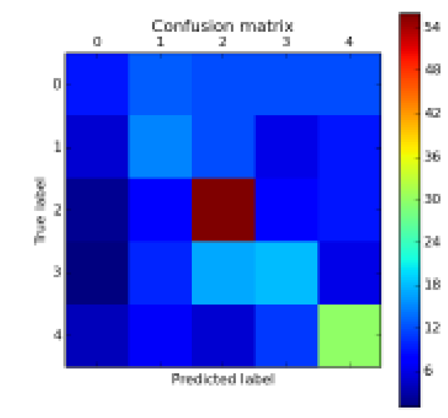
Classifier	Accuracy (w/o PCA)	Accuracy (with PCA)	Precision	Recall	F1-score
SVM (rbf)	37.77%	43.33%	0.52	0.43	0.45
Naive Bayes	36.6%	39.73%	0.41	0.40	0.40
QDA	37.03%	37.37%	0.39	0.37	0.37
LDA	37.7%	41.07%	0.44	0.41	0.42
Ada Boost	30.97%	36.02%	0.37	0.36	0.36
k-NN	33.33%	38.88%	0.41	0.39	0.40
Log-Regression	39.05%	42.42%	0.48	0.42	0.45
Decision tree	27-32%	27-32%	0.28	0.28	0.27
Random Forest	32-38%	32-38%	0.36	0.35	0.35
Ensemble	42.76%	41.41%	0.48	0.43	0.44



(a) SVM (b) LDA (c) Logistic Regression

(d) QDA (e) Naive Bayes (f) k-NN

(g) Ada Boost (h) Random Forest (i) Decision Tree



Confusion matrix for Ensemble Classifier

Conclusion:

- ➢ The ensemble classier used hard voting technique to classify music clips. We were able to achieve better precision using the ensemble classier.
- ➢ We were also able to establish the fact that increasing the number of features does no necessarily increase linearly the accuracy of prediction results.
- ➢ Support Vector Machine with rbf kernel, gamma = 0.01 and C = 10, and Ensemble Classier were the best predictors.
- ➢ Use of PCA with 14 and 22 components respectively, resulted into the mentioned results.
- ➢ According to the researcher Renato Panda, this dataset is might be more difficult than the MIREX dataset.
- ➢ Further improvements in the results could have been improved by using better methods of splitting into training and test datasets and using ReliefF feature selection technique.

References:

- ❖ Renato Panda and Rui Pedro Paiva. Music emotion classication: Dataset acquisition and comparative analysis. In 15th International Conference on Digital Audio Eects (DAFx-12). Citeseer, 2012.
- ❖ Renato Panda, Bruno Rocha, and Rui Pedro Paiva. Music emotion recognition with standard and melodic audio features. Applied Articial Intelligence, 29(4):313{334, 2015.
- ❖ R Shobana. A framework for audio feature extraction from videos for genre identication. 2015.
- ❖ Alicja Wieczorkowska, Piotr Synak, and Zbigniew W Ras. Multi-label classication of emotions in music. In Intelligent Information Processing and Web Mining, pages 307{315. Springer, 2006.