HumoRadio

A music Recommender system based on user's mood

Business Intelligence Course

Daniele Battista
Tommaso Vongher

A.Y. 2014/2015

RoadMap Overview

1. *Idea*
2. *Technologies Used*
3. *Dataset Collection*
4. *How the Application works*
5. *Running Demonstration*
6. *Clustering Issues*
7. *Conclusions*

Idea

Our idea is to create a recommender system based on user's mood, with 2 fields: Happy/Sad and Relaxing/Exciting Levels.

Step 1	*Step 2*	*Step 3*
The user selects the music genre and provides his mood status.	*The application selects a collection of songs to be listened based on an apriori made survey.*	*The user implicitly rates this songs for further selections.*

Technologies Used

We developed a Java web based application using JEE. The application consists of an HTML page in which the user specifies a genre, if he wants, and a specific mood or he leaves it indifferent. In another HTML5 page a servlet will manage requests coming from the first page and will provide user the playlist to listen or will let the user come back to change his mood.

The application will play songs through Youtube, displaying also the video related in order to have a better entertainment. We worked with the Youtube Api to interact with player parameters.

We exploited Java to manage all the logic behind the service, Javascript to handle Youtube player aspects and user interaction.

NetBeans IDE with Glassfish server were used as environment of the project.

Technologies Used

NetBeans

GlassFish

Java EE™

CSS

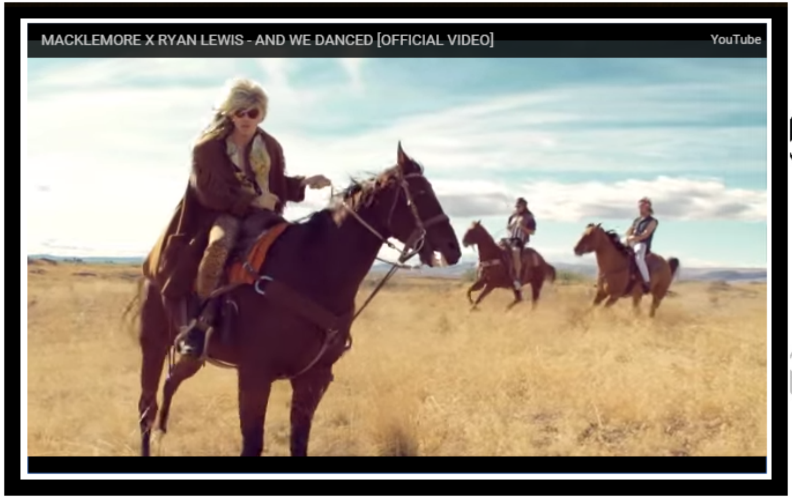
HTML

JavaScript

YouTube

WEKA
The University
of Waikato

Overview



CHANGE YOUR CHOICE!

And we danced – Macklemore

Dataset Collection

Cold start problem:

- *online survey on Google form to*

Dataset structure:

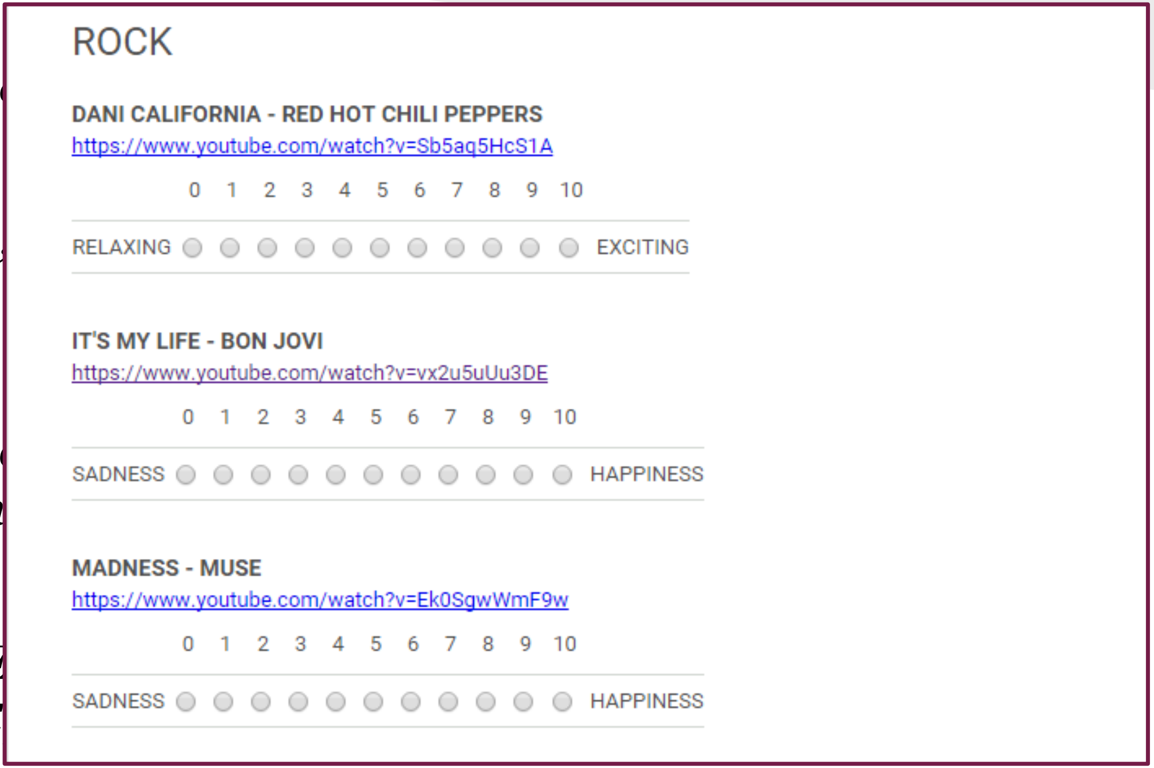
- ***5** genres (**Rock**, **Pop**, **Blues***
- ***30** songs per genre*

Rating values:

- ***2** mutual exclusive bar of mo*
- ***10** degrees of freedom withi*

Tables created:

1. ***SONG_DATA(TITLE**, TIME*
2. ***SONG_RATING(ID**, TITLE*


ROCK

DANI CALIFORNIA - RED HOT CHILI PEPPERS
https://www.youtube.com/watch?v=Sb5aq5HcS1A
0 1 2 3 4 5 6 7 8 9 10
RELAXING EXCITING

IT'S MY LIFE - BON JOVI
https://www.youtube.com/watch?v=vx2u5uUu3DE
0 1 2 3 4 5 6 7 8 9 10
SADNESS HAPPINESS

MADNESS - MUSE
https://www.youtube.com/watch?v=Ek0SgwWmF9w
0 1 2 3 4 5 6 7 8 9 10
SADNESS HAPPINESS

Dataset Collection

#	TITLE	TIME	LINK
1	Pretty Fly – Offspring	192	nzY2Qcu5i2A
2	Basket case – Green Day	192	NUTGr5t3MoY
3	Knocking on Heaven's door – Guns n' Roses	341	2tmc8rJgxUI
4	Sweet Home Alabama - Lynyrd Skynyrd	300	ye5BuYf8q4o
5	Roxanne – The Police	314	3T1c7GkzRQQ
6	Dani California – Red Hot Chili Peppers	328	Sb5aq5HcS1A
7	Wonderwall – Oasis	318	bx1Bh8ZvH84
8	Do I wanna know – Artic Monkeys	305	bpOSxM0rNPM
9	Seven nation army – The White Stripes	240	0J2QdDbelmY
10	Zombie – Cranberries	316	6Ejga4kJUts
11	21 guns – Green Day	327	r00ikilDxW4
12	Can't stop – Red Hot Chili Peppers	278	BfOdWSiyWoc

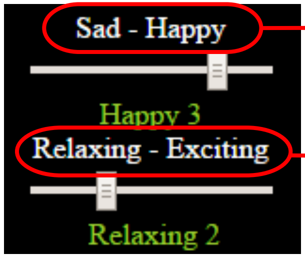
SONG_DATA

SONG_RATING

#	ID	TITLE	RE1	RE2	RE3	RE4	RE5	RE6	RE7	RE8	RE9	RE10	SH1	SH2	SH3	SH4	SH5	SH6	SH7	SH8	SH9	SH10	GENRE
1	nzY2Qcu5i2A	Pretty fly – Offspring	1	1	0	1	1	1	1	8	2	6	1	0	2	4	0	0	3	5	4	3	1
2	NUTGr5t3MoY	Basket case – Green Day	1	0	0	1	2	3	4	3	5	3	1	0	1	1	1	4	4	4	4	2	1
3	2tmc8rJgxUI	Knocking on Heaven's door – Guns ...	4	2	2	5	3	1	0	3	0	4	2	0	3	0	2	3	5	2	2	3	1
4	ye5BuYf8q4o	Sweet home Alabama - Lynyrd Sky...	2	3	1	2	0	2	3	2	5	2	0	1	0	0	1	1	4	6	5	3	1
5	3T1c7GkzRQQ	Roxanne – The Police	2	1	3	5	1	3	1	2	1	0	1	2	1	2	2	2	4	0	1	0	1
6	Sb5aq5HcS1A	Dani California – Red Hot Chili Pepp...	2	0	0	3	0	4	5	3	3	2	1	0	0	3	2	2	4	5	2	2	1
7	bx1Bh8ZvH84	Wonderwall – Oasis	6	2	3	6	2	1	2	2	0	1	3	2	1	2	3	1	1	3	3	1	1
8	bpOSxM0rNPM	Do I wanna know – Artic Monkeys	3	0	3	3	0	3	4	1	0	0	2	3	1	3	4	2	0	1	0	0	1
9	0J2QdDbelmY	Seven nation army – The White Str...	1	0	1	1	0	0	3	4	1	8	1	0	0	2	1	1	3	5	1	4	1

How the Application works

Creates a decreasing rank based on the sum of the support each song has in the fields selected.



*Fields **SH1-SH10** in Dataset*

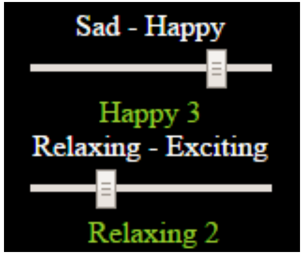
*Fields **RE1-RE10** in Dataset*

<u>Possible combinations of selections</u>:

1. *2 values not “Indifferent”*
2. *1 value “Indifferent” and the other not*
3. *Both “Indifferent”*

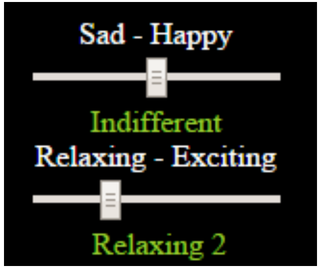
*When an “**Indifferent**” value is selected we perform the sum on the fields right above and below the middle one (which are SH5, SH6 and RE5, RE6).*

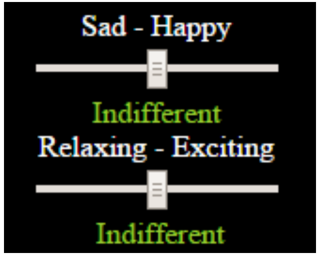
How the Application works



In this case the application sums the related 2 values in order to get the rank

In this case one values has been left “Indifferent” and, for this reason, the application sums the specific value for the selected aspect of mood and 2 values that are the 2 boundary values for the other aspect.
In this example we take for the Sad-Happy field the values Sad = 5 and Happy = 1 which are the next boundaries of the middle "Indifferent" value.



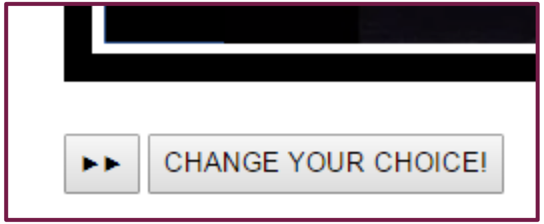


In this last case both are left “Indifferent” and so the procedure here is repeated for both aspects, considering 4 values for the final sum.

How the Application works

The application creates 3 ranks of 10 songs divided as shown in the picture below. After a complete rank has been listened, the application asks the user if he wants to change his mood or continue listening songs maintaining the previous mood.

However, if the user wants, he can change his mood whenever he wants. At the end of the third rank, the application comes back automatically at the selection page to choose a new mood.



The user is not able to change the current execution time of a song in order not to alter statistics in the dataset. ***How is the dataset updated?***

How the Application works

The updating task is performed implicitly taking into consideration the percentage of listened time (t) w.r.t. the total song time. When a song is proposed to the user, the application considers significant the moment at which he skips the song and it adds a value to the current rate present in the dataset.
If a song is listened entirely, the last case will be triggered.

+0

Percentage: ***t <= 25%***
The user consider this song ***not appropriate*** *for this mood.*
No points is added to the dataset rate

+1

Percentage: ***25% < t <= 50%***
The user consider this song ***not so appropriate*** *for this mood.*
One point is added to the dataset rate

+2

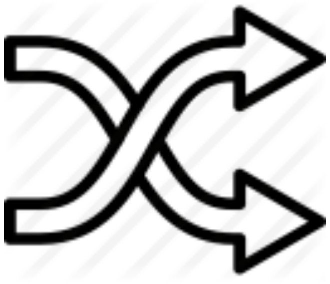
Percentage: ***50% < t <= 75%***
The user consider this song ***appropriate*** *for this mood.*
Two points are added to the dataset rate

+3

Percentage: ***t > 75%***
The user consider this song ***very appropriate*** *for this mood.*
Three points are added to the dataset rate

How the Application works

How to avoid the situation in which the user, selecting twice the same mood, has to play the same selection of songs?



*After songs have been selected and all three ranks have been set up, on these three selections a **shuffle function** is applied to mix the songs disposal.*

Random Selection

Clustering Issues

A-posteriori Clustering to visualize similar songs that represent a specific mood.

- *Data Preprocessing*
- *Hierarchical Clustering*
- *K-Means Clustering*
- *DBSCAN Clustering*

We were rates had through s could hav th

This lead u

We use data hierar exploitin

We dendrogr "similarit the cut numb

Now we value cluster Means evaluat result, c that it is

To free ourselves from this input parameter, we decided to apply the DBSCAN algorithm on the dataset, trying to find best values for ε and MinPTs, taking care of of outliers

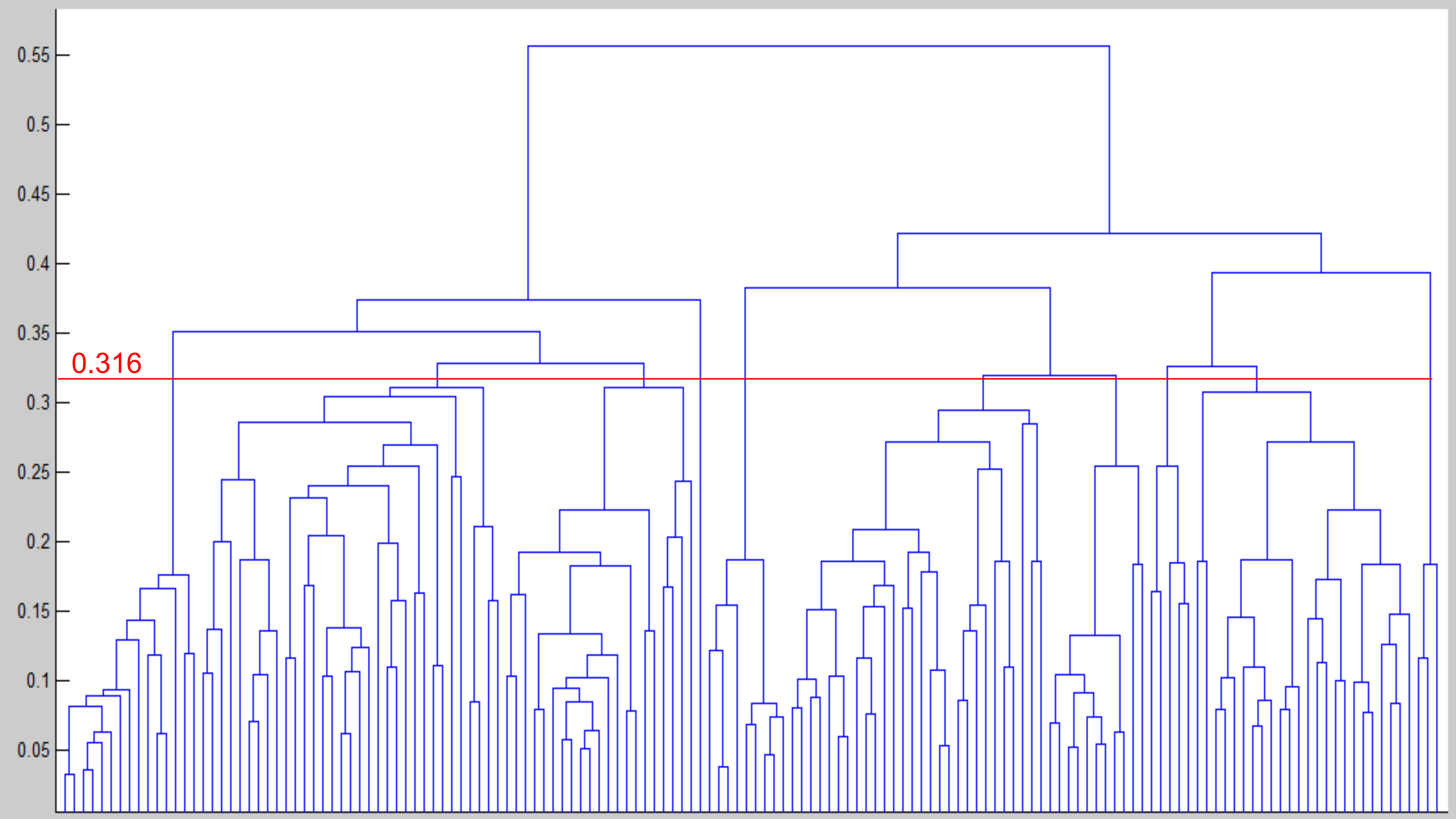
Data Preprocessing

Our database values can be considered as counters which could be increased at each selection.
We could have songs more listened than others.
In order to have comparable weights on every song we decided to ***Normalize*** *the entire dataset.*

We used the Unsupervised filter Normalize provided by Weka Api.

14	vx2u5uUu3DE	It's my life – Bon Jovi	0	1	0	0	2	3	0	3	4	13	1	0	1	1	0	4	2	6	2	10	1
20	L_jWHffIx5E	All star – Smash Mouth	0	2	0	2	0	1	4	2	3	3	1	0	1	0	1	4	3	2	4	3	1

Once the dataset is normalized, we are allowed to use the Euclidean Distance to compute the similarities between objects.


0.55
0.5
0.45
0.4
0.35
0.316
0.3
0.25
0.2
0.15
0.1
0.05

K-Means Clustering

kMeans
======

Number of iterations: 6
Within cluster sum of squared errors: 81.37120804631384
Missing values globally replaced with mean/mode

Cluster centroids:

		Cluster#									
Attribute	Full Data	0	1	2	3	4	5	6	7		
	(150)	(14)	(12)	(30)	(24)	(10)	(13)	(5)	(15)		
RE1	0.1798	0.1664	0.1768	0.0438	0.0627	0.2501	0.2093	0.3696	0.5706		
RE2	0.1103	0.0453	0.1438	0.0315	0.0362	0.262	0.1897	0.2791	0.2187		
RE3	0.1999	0.3689	0.1483	0.1238	0.0782	0.5136	0.3374	0.39	0.211		
RE4	0.1978	0.2368	0.2017	0.1508	0.1093	0.1438	0.3623	0.175	0.2534		
RE5	0.1593	0.2427	0.1819	0.1759	0.089	0.1795	0.1699	0.0969	0.0858	0.2621	0.0852
RE6	0.1788	0.1521	0.1593	0.3291	0.1622	0.0682	0.131	0.1547	0.0463	0.25	0.1051
RE7	0.2186	0.1906	0.3418	0.3711	0.2925	0.0881	0.099	0.0229	0.0684	0.1445	0.204
RE8	0.1676	0.1205	0.186	0.2437	0.3257	0.0497	0.0638	0.0775	0.0542	0.0754	0.2144
RE9	0.0887	0.0251	0.047	0.0531	0.303	0	0.0217	0.0229	0.0086	0.0761	0.1618
RE10	0.1208	0.0706	0.1476	0.0842	0.2057	0	0.0122	0.1255	0.0048	0.0619	0.5505
SH1	0.1541	0.137	0.1756	0.0686	0.0726	0.2106	0.1953	0.2084	0.4569	0.1252	0.0684
SH2	0.0848	0.0178	0.2385	0.034	0.0233	0.1154	0.1728	0.3758	0.089	0.0881	0
SH3	0.152	0.1356	0.3084	0.0813	0.0532	0.1769	0.4107	0.1976	0.1795	0.1416	0.0387
SH4	0.2313	0.5298	0.3265	0.0971	0.0781	0.4748	0.3447	0.2184	0.3004	0.197	0.0545
SH5	0.184	0.1468	0.2381	0.2836	0.0893	0.1492	0.1795	0.1397	0.1194	0.2885	0.1006
SH6	0.1954	0.1285	0.176	0.3053	0.2091	0.1165	0.1679	0.0636	0.0807	0.33	0.0976
SH7	0.2274	0.1472	0.1541	0.3236	0.3841	0.0929	0.1515	0.1441	0.0833	0.26	0.2045
SH8	0.1668	0.1331	0.0693	0.2137	0.347	0.0758	0.0347	0.2346	0.0606	0.0933	0.2544
SH9	0.0919	0.0825	0.068	0.1113	0.1966	0.0332	0	0.0763	0.035	0.0417	0.1678
SH10	0.1007	0.0052	0.046	0.0772	0.1783	0.009	0.0355	0.053	0.0136	0.081	0.5057

Clustered Instances

0	14	(9%)
1	12	(8%)
2	30	(20%)
3	24	(16%)
4	10	(7%)
5	13	(9%)
6	5	(3%)
7	15	(10%)
8	16	(11%)
9	11	(7%)

K-Means Clustering

```
63->    3,1,7,2,4,2,3,0,0,0,1,0,4,8,1,2,1,3,0,1->     0->    Me and the devil blues – Eric Clapton
73->    1,0,2,4,1,2,3,1,1,1,1,0,1,3,2,1,1,3,2,0->     0->    Sweet home Chicago – Eric Clapton
74->    1,0,2,2,2,1,3,1,0,1,1,0,0,4,0,1,3,1,0,0->     0->    Mannish boy – Muddy Waters
77->    1,0,3,4,1,3,0,3,0,1,1,0,2,3,1,1,2,2,0,0->     0->    Dust my broom – Elmore Jones
78->    2,0,10,1,3,1,2,2,0,0,1,0,1,10,1,3,2,2,0,0->   0->    My babe – Little Walter
```

```
6->     8,2,3,6,2,1,2,2,0,1,5,2,1,2,3,1,1,3,3,1->    7->    Wonderwall – Oasis
29->    8,2,4,4,1,0,1,3,0,0,4,0,4,9,0,0,1,2,1,0->    7->    Hotel california – Eagles
38->    7,2,3,1,3,0,4,0,1,0,7,2,4,5,3,1,1,2,0,0->    7->    I will always love you – Whitney Houston
39->    12,4,6,4,1,1,0,0,0,0,13,1,5,4,1,1,1,0,0,1->  7->    My heart will go on – Celine Dion
40->    12,4,4,4,0,0,1,1,1,0,8,0,5,3,1,2,0,2,0,0->   7->    Yesterday – Beatles
90->    6,2,4,2,1,1,1,4,0,0,6,1,1,4,2,1,1,2,0,1->    7->    Con te partirò – Bocelli
96->    14,1,1,4,0,0,0,0,0,0,9,2,3,4,2,2,1,0,0,0->   7->    Any other name – Thomas Newman
101->   6,3,3,5,2,1,1,1,0,0,4,0,2,3,1,2,6,1,1,0->    7->    Now we are free – Enya
103->   13,5,4,1,0,1,1,0,0,0,7,3,2,5,1,1,1,1,1,0->   7->    Rivers flow in you – Yiruma
104->   5,7,1,4,1,0,0,0,0,0,5,2,1,6,2,1,1,0,0,0->    7->    May it be – Enya
107->   5,4,2,3,1,1,1,0,0,0,4,1,1,3,2,1,0,0,1,0->    7->    Lord of the rings – Piano Guys
109->   7,2,3,5,1,1,0,0,0,0,5,1,1,5,1,1,1,0,0,0->    7->    Only time – Enya
111->   10,1,1,5,1,0,1,0,0,0,10,2,3,3,1,1,0,0,0,0->  7->    Love me – Yiruma
113->   7,3,3,3,2,1,0,0,0,0,6,1,5,3,2,1,2,0,0,0->    7->    Nightbook – Ludovico Einaudi
117->   6,3,3,2,1,1,1,0,0,0,7,1,2,4,2,1,0,0,0,0->    7->    Time – Hans Zimmer
```

```
28->    1,0,0,1,1,0,2,3,5,14,1,0,0,1,2,1,1,3,3,14->   9->    Don't stop me now – Queen
30->    0,0,1,2,1,2,10,6,4,8,1,0,0,1,1,0,12,5,3,13->  9->    Happy – Pharrel Williams
47->    1,0,0,0,2,0,7,7,5,9,1,0,0,0,1,0,8,6,4,11->    9->    Mamma mia – Abba
50->    1,0,0,1,2,2,1,1,3,6,1,0,0,1,1,1,4,4,3,3->     9->    Thriller – Michael Jackson
133->   0,0,0,2,1,2,3,3,2,4,1,0,1,0,2,3,0,4,3,4->     9->    Yeah! - Usher
```

RE 3-4
SH 4-5

RE 1
SH 1

RE 10
SH 10

DBSCAN Clustering

```
DBSCAN clustering results
========================================================================================

Clustered DataObjects: 150
Number of attributes: 20
Epsilon: 0.7; minPoints: 2
Index: weka.clusterers.forOPTICSAndDBScan.Databases.SequentialDatabase
Distance-type: weka.clusterers.forOPTICSAndDBScan.DataObjects.EuclideanDataObject
Number of generated clusters: 11
Elapsed time: ,06

(  0.) 0.065653,0.065653,0,0.065653,0.131306,0.065653,0.131306,0.459573,0.13  -->  0
(  1.) 0.082761,0,0,0.082761,0.165521,0.248282,0.331042,0.248282,0.413803,0.  -->  1
(  2.) 0.257248,0.171499,0.171499,0.428746,0.257248,0.085749,0,0.257248,0,0.  -->  NOISE
(  3.) 0.16169,0.242536,0.080845,0.16169,0,0.16169,0.242536,0.16169,0.404226  -->  NOISE
(  4.) 0.210819,0.105409,0.316228,0.527046,0.105409,0.316228,0.105409,0.2108  -->  NOISE
(  5.) 0.167248,0,0,0.250873,0,0.334497,0.418121,0.250873,0.250873,0.167248,  -->  7
(  6.) 0.57886,0.144715,0.217072,0.434145,0.144715,0.072357,0.144715,0.14471  -->  NOISE
(  7.) 0.304604,0,0.304604,0.304604,0,0.304604,0.406138,0.101535,0,0,0.20306  -->  NOISE
(  8.) 0.050965,0,0.050965,0.050965,0.152894,0.152894,0.152894,0.203859,0.05  -->  0
(  9.) 0.228748,0.304997,0,0.228748,0.152499,0.076249,0.381246,0.228748,0.07  -->  NOISE
( 10.) 0.162221,0,0.324443,0.081111,0.243332,0.243332,0.243332,0.324443,0,0,  -->  NOISE
( 11.) 0.074125,0,0.074125,0.074125,0.074125,0.074125,0.44475,0.074125,0.444  -->  1
( 12.) 0,0,0.178885,0.178885,0,0.268328,0.447214,0.178885,0.089443,0.447214,  -->  NOISE
( 13.) 0,0.043561,0,0,0.087121,0.130682,0,0.130682,0.174243,0.696971,0.04356  -->  0
( 14.) 0.076923,0,0,0.153846,0,0.230769,0.153846,0.461538,0.384615,0.230769,  -->  NOISE
( 15.) 0.094281,0.04714,0.04714,0,0.094281,0.094281,0,0.141421,0.04714,0.612  -->  0
( 16.) 0.079556,0,0.079556,0,0.079556,0.079556,0.397779,0.318223,0.318223,0.  -->  1
( 17.) 0,0,0.128831,0.128831,0.322078,0.386494,0.386494,0.193247,0,0.064416,  -->  5
```

```
Clustered Instances

 0       11 ( 14%)
 1       16 ( 20%)
 2        3 (  4%)
 3       12 ( 15%)
 4        2 (  2%)
 5       15 ( 19%)
 6        2 (  2%)
 7        2 (  2%)
 8        4 (  5%)
 9       12 ( 15%)
10        2 (  2%)

Unclustered instances : 69
```

The value of ε and MinPts were chosen based on different trials

DBSCAN Clustering

```
0->     1,1,0,1,2,1,2,7,2,8,1,0,2,2,1,0,4,5,4,6->        0.0->   Pretty Fly – Offspring
8->     1,0,1,1,3,3,3,4,1,13,1,0,0,1,4,4,3,5,1,10->      0.0->   Seven nation army – The White Stripes
13->    0,1,0,0,2,3,0,3,4,16,1,0,1,1,0,4,2,6,2,13->      0.0->   It's my life – Bon Jovi
15->    2,1,1,0,2,2,0,3,1,13,3,0,1,0,4,0,2,1,1,15->      0.0->   Bhoemian rapsody – Queen
20->    1,0,1,1,0,2,6,2,1,12,1,0,2,1,2,1,4,3,3,11->      0.0->   Sweet child of mine – Guns N'Roses
24->    1,0,1,0,0,2,6,3,3,8,1,0,0,2,1,3,2,5,3,6->        0.0->   Eye of thr tiger – Survivor
```

```
36->    3,3,6,4,1,1,0,1,0,0,1,0,5,6,2,3,1,1,0,0->     9.0->   Angels – Robbie Williams
52->    3,3,8,1,1,1,0,1,0,0,1,2,1,4,2,2,2,1,0,0->     9.0->   She's the one – Robbie Williams
63->    3,1,7,2,4,2,3,0,0,0,1,0,4,8,1,2,1,3,0,1->     9.0->   Me and the devil blues – Eric Clapton
70->    4,2,10,3,3,0,2,0,0,0,2,0,4,7,4,1,0,0,0,0->    9.0->   You got the silver – Susan Tedeschi
71->    3,2,10,3,2,0,0,1,0,0,2,2,3,9,1,1,0,0,1,0->    9.0->   The sky is crying – Elmor Jones
78->    2,0,10,1,3,1,2,2,0,0,1,0,1,10,1,3,2,2,0,0->   9.0->   My babe – Little Walter
80->    1,1,8,6,0,1,2,0,0,0,1,0,1,8,1,2,2,0,0,0->     9.0->   Turner station – Eric Bibb
83->    4,0,8,2,3,1,1,1,0,0,1,0,2,10,3,0,1,0,2,0->    9.0->   Help me – Sonny Boy Williamson
87->    2,0,7,3,2,1,1,1,0,0,2,0,2,9,0,1,2,0,0,0->     9.0->   Spoonful – Howlin' Wolf
88->    2,2,6,1,1,1,1,1,0,0,2,1,0,5,1,0,1,2,1,0->     9.0->   I'm your hocchi coochie man – Muddy Waters
89->    4,0,7,3,1,0,1,0,0,0,2,0,2,8,5,3,2,1,0,0->     9.0->   Stop! - Joe Bonamassa
91->    2,1,6,2,3,1,1,0,0,0,3,1,4,3,1,1,1,0,0,1->     9.0->   Melodramma – Bocelli
```

```
103->   13,5,4,1,0,1,1,0,0,0,7,3,2,5,1,1,1,1,1,0->     3.0->   Rivers flow in you – Yiruma
104->   5,7,1,4,1,0,0,0,0,0,5,2,1,6,2,1,1,0,0,0->      3.0->   May it be – Enya
107->   5,4,2,3,1,1,1,0,0,0,4,1,1,3,2,1,0,0,1,0->      3.0->   Lord of the rings – Piano Guys
109->   7,2,3,5,1,1,0,0,0,0,5,1,1,5,1,1,1,0,0,0->      3.0->   Only time – Enya
111->   10,1,1,5,1,0,1,0,0,0,10,2,3,3,1,1,0,0,0,0->    3.0->   Love me – Yiruma
113->   7,3,3,3,2,1,0,0,0,0,6,1,5,3,2,1,2,0,0,0->      3.0->   Nightbook – Ludovico Einaudi
117->   6,3,3,2,1,1,1,0,0,0,7,1,2,4,2,1,0,0,0,0->      3.0->   Time – Hans Zimmer
```

RE 10
SH 10

RE 3-4
SH 4-5

RE 1
SH 1

Thanks

"The validation of clustering structures is the most difficult and frustrating part of cluster analysis.

Without a strong effort in this direction, cluster analysis will remain a black art accessible only to those true believers who have experience and great courage."

Algorithms for Clustering Data, Jain and Dubes