



UNIVERSIDADE D  
**COIMBRA**

Multimédia

TP2 - Music Information Retrieval

Dinis Isaev N°2021212927  
Marcelo Gomes N°2021222994  
Pedro Brites N°2021226319  
Carlos Ferreira N°2021226028

# Índice

<b>Métricas de Similaridade.....</b>	<b>2</b>
<b>Avaliação objetiva.....</b>	<b>5</b>
<b>Avaliação subjetiva.....</b>	<b>6</b>

# Métricas de Similaridade

De seguida apresentamos as tabelas com os resultados obtidos:

## Ranking Euclideano

1	MT0003949060	2.07155069
2	MT0004274911	2.09657571
3	MT0001515531	2.12421922
4	MT0003900455	2.15217283
5	MT0009897495	2.1572501
6	MT0000040632	2.16166885
7	MT0004032071	2.16881018
8	MT0005469880	2.17979582
9	MT0007043936	2.20076728
10	MT0034005433	2.21911

### Ranking de Manhattan

1	MT0003949060	15.73271729
2	MT0004274911	17.26874492
3	MT0000040632	17.71159109
4	MT0000218346	18.00354683
5	MT0003900455	18.05985231
6	MT0005469880	18.20241476
7	MT0008401073	18.32846813
8	MT0001515531	18.44261891
9	MT0001624303	18.51657708
10	MT0034125967	18.52620144

### Ranking do Cosseno

1	MT0004274911	0.05808815
2	MT0003949060	0.05873648
3	MT0001515531	0.06169296
4	MT0002634024	0.06210557
5	MT0003900455	0.06282674
6	MT0009897495	0.06313855
7	MT0001942272	0.0632992
8	MT0004032071	0.06342195
9	MT0000040632	0.06381533
10	MT0005752234	0.06484468

Ao observar os resultados, percebemos que as posições das músicas no top10 variam conforme a métrica utilizada. Assim, podemos concluir que cada métrica de similaridade tem formas específicas de calcular a distância entre os dados. Essas diferenças podem levar a variações nos resultados.

A distância euclidiana, por exemplo, enfatiza a proximidade direta no espaço de características, enquanto a distância de Manhattan destaca a soma das diferenças absolutas entre as características. Por outro lado, a distância do cosseno mede a similaridade com base na direção dos vetores de características, sem considerar sua magnitude. Essas diferenças nas abordagens de cálculo de similaridade têm implicações significativas na identificação das músicas mais similares à query.

## Avaliação objetiva

Para uma avaliação objetiva, extraímos os metadados de todas as músicas do ficheiro *panda\_dataset\_taffc\_metadata.csv* e de acordo com os dados da query, atribuímos pontos por semelhanças nas categorias Artist, MoodsStrSplit e GenresStr. Com isto obtivemos o seguinte ranking:

Música	Nome	Pontos de semelhança
1	MT0000040632	7
2	MT0033397838	7
3	MT0002222957	7
4	MT0003949060	7
5	MT0010487769	7
6	MT0010489498	7
7	MT0012331779	7
8	MT0027048677	7
9	MT0008222676	3
10	MT0007840454	3

Com estes valores, agora calculamos a variável **Precisão**, que indica qual a percentagem de músicas do ranking de cada matriz de similaridade (euclidiana, manhattan e cosseno), se encontra no ranking baseado na comparação dos metadados. Neste caso, resultaram os seguintes valores de precisão:

- Precision de: 20%
- Precision dm: 20%
- Precision dc: 20%

Com estes valores de precisão tão baixos conseguimos perceber que é bastante diferente a maneira como os rankings (metadados e distâncias) são calculados.

# Avaliação subjetiva

Abaixo estão apresentadas as tabelas com a avaliação subjetiva para cada um dos rankings construídos usando as 3 métricas de distâncias assim como para o ranking de similaridade com base nos metadados. No final, em cada ranking, foi-se também calculada a média e o desvio padrão, para cada uma das músicas, e a média e o desvio padrão global.

## Ranking Euclideano

Música	1	2	3	4	5	6	7	8	9	10
Carlos	4	2	1	1	2	5	3	2	2	1
Dinis	4	2	2	1	1	5	3	1	3	1
Marcelo	4	1	2	1	1	4	2	1	3	1
Pedro	5	2	1	1	1	5	2	2	3	1
Média	4.25	1.75	1.5	1	1.25	4.75	2.5	1.5	2.75	1
Desvio Padrão	0.433	0.433	0.5	0	0.433	0.433	0.5	0.5	0.433	0

Média total: 2.225

Desvio padrão total: 1.332

Precision: 0.4

## Ranking de Manhattan

Música	1	2	3	4	5	6	7	8	9	10
Carlos	4	2	4	2	1	3	3	1	1	2
Dinis	4	2	5	2	1	1	2	2	1	1
Marcelo	4	2	4	1	1	1	2	1	1	2
Pedro	5	2	5	2	1	2	2	1	1	2
Média	4.75	2	4.5	1.75	1	1.75	2.75	1.75	1	1.75
Desvio Padrão	0.433	0	0.5	0.433	0	0.829	0.433	0.433	0	0.433

Média total: 2.15

Desvio padrão total: 1.256

Precision: 0.3

### Ranking do Cosseno

Música	1	2	3	4	5	6	7	8	9	10
Carlos	2	5	1	2	1	2	1	2	4	1
Dinis	2	5	2	1	1	1	1	1	5	1
Marcelo	2	5	1	1	1	1	1	2	5	1
Pedro	2	5	1	2	1	2	1	2	4	1
Média	2	5	1.25	1.5	1	1.5	1	1.75	4.5	1
Desvio Padrão	0	0	0.433	0.5	0	0.5	0	0.433	0.5	0

Média total: 2.05

Desvio padrão total: 1.431

Precision: 0.2

### Ranking conjunto

Música	1	2	3	4	5	6	7	8	9	10
Carlos	4	5	5	5	4	5	5	5	2	1
Dinis	5	5	5	4	3	5	5	5	3	1
Marcelo	5	4	5	4	3	5	5	5	3	1
Pedro	4	5	5	4	4	5	5	5	2	1
Média	4.5	4.75	5	4.25	3.5	5	5	5	2.5	1
Desvio Padrão	0.5	0.433	0	0.433	0.5	0	0	0	0.5	0

Média total: 4.05

Desvio padrão total: 1.322

Precision: 0.9

Ao observar as tabelas, notamos que, das 3 métricas, o ranking Euclideano, apesar de não ser muito certo, tendo, por isso uma média de avaliação baixa, foi o mais preciso, seguido do ranking de Manhattan e, o menos preciso, foi o ranking do Cosseno. No entanto, apesar dos 3 apresentarem uma precisão extremamente baixa, o ranking de similaridade com base nos metadados, foi extremamente preciso, sendo quase todas as músicas parecidas à query.