

# L'ADN dels èxits musicals: Anàlisi i recomanació basada en patrons

Marc Puigbó Paricio

## 1 INTRODUCCIÓ - CONTEXT DEL TREBALL

EN els darreres anys, les plataformes digitals han facilitat l'accés a la música, la qual cosa ha permès escoltar a baix cost tota mena de cançons sense limitacions, i descobrir una gran varietat de música mitjançant sistemes de recomanació. Fet que ha provocat la constant aparició de noves tendències musicals, i amb elles milers de cançons d'èxit, que porta a qüestionar-se si existeix algun factor comú que contribueix a l'èxit de les cançons.

Les recomanacions solen tenir similituds en la popularitat, l'artista o inclús la presència de les cançons en les llistes de reproducció més escoltades, que acostumen a ser influenciades per les discogràfiques [1]. Com a resultat, la recomanació pot presentar un biaix que tendeix a afavorir els artistes consolidats, i a deixar en un segon pla les propostes emergents o menys conegudes.

Aquest treball de fi de grau s'emmarca dins d'un projecte en col·laboració amb la meua companya Lucía Carmona, que busca detectar l'existència de patrons entre els èxits musicals del moment, i com poden ser utilitzats per generar cançons amb un gran potencial. Concretament, aquest treball se centra en la fase prèvia d'exploració i anàlisi de similituds entre les cançons més escoltades de l'any passat, i en com desenvolupar un sistema de recomanació basat en el contingut acústic de les cançons.

La hipòtesi que es planteja, és que els sistemes de recomanació actuals prioritzen criteris comercials per sobre del contingut musical, cosa que pot limitar el descobriment de cançons, que a pesar de compartir similituds amb les cançons del moment, manquen de visibilitat. Mitjançant aquesta anàlisi de similitud acústica, es busca aportar una alternativa més objectiva en la recomanació de cançons.

Per tal d'aconseguir la finalitat del projecte, es plantejen els següents objectius:

- Detectar i analitzar patrons sonors entre cançons a partir de les seves característiques acústiques i estructurals, explorant diferents mètodes de processament de senyal, i aprenentatge computacional.
- Desenvolupar un sistema de recomanació basat en contingut, considerant les característiques acústiques

de les cançons, i les seves metadades, evitant factors externs com la popularitat o el nombre de reproduccions.

Adicionalment, es planteja un objectiu secundari interessant a realitzar si es té el temps i els recursos necessaris per portar-lo endavant, es vol fer un experiment amb un electroencefalograma (EEG) a un conjunt de persones, per analitzar les seves reaccions cerebrals a cançons similars, reforçant així l'anàlisi dels patrons trobats i la funcionalitat del sistema de recomanació.

## 2 PLANIFICACIÓ

Pel desenvolupament del treball, es planteja una planificació estructurada en diferents fases definides, que facilita una millor organització i comprensió del procés.

En primer lloc, es crearà la base de dades que contingui les cançons més escoltades de l'any passat, es farà una recerca per saber d'on obtenir les dades, quins atributs de les dades es necessiten, i quin processament dur a terme.

En segon lloc, es farà una exploració sobre els diferents mètodes d'extracció de característiques que hi ha, quins són els més representatius, els que més s'utilitzen en projectes semblants...

En tercer lloc, cal estudiar tècniques per detectar patrons similars entre cançons fent servir les extraccions de característiques. S'analitzarà el comportament de les tècniques usades en articles amb objectius similars.

Finalment, es farà un estudi sobre l'estat de l'art dels sistemes de recomanació, s'analitzarà els sistemes de les plataformes digitals més utilitzades, i d'articles existents, per desenvolupar un sistema de recomanació.

## 3 ESTAT DE L'ART

En aquesta secció es detalla l'estat de l'art, una part important per entendre el contingut teòric del projecte, i el marc actual dels sistemes que volem desenvolupar.

### 3.1 Extracció de característiques

L'extracció de característiques és una part necessària per detectar similituds entre cançons, és la base per obtenir bons resultats. És una fase present en la gran majoria d'articles sobre recuperació d'informació musical (MIR).

Diversos estudis utilitzen tècniques d'extracció aplicades a diferents dominis. L'article escrit per Mahajan, Kaavya (2023) [2], fa servir mètriques com *chroma\_stft* (ener-

- 
- E-mail de contacte: marc.puigbo@autonoma.cat
  - Treball tutoritzat per: Debora Gil Resina i Aura Hernandez Sabate (Àrea de Ciències de la Computació i Intel·ligència Artificial)
  - Curs 2024/25

gia harmònica distribuïda per notes musicals), *rmse* (nivell d'energia del senyal), *spectral\_centroid* (centre de gravetat espectral que indicia la brillantor del so) i, *spectral\_rolloff* (freqüència per sota de la qual es concentra la major part de l'energia espectral), amb l'objectiu de classificar cançons. La investigació de Méndez, Miguel (2020) [3], es basa en l'ús de l'espectrograma, aplicant filtres per extreure una empremta digital única de cada fragment d'àudio, orientat al reconeixement automàtic musical. L'estudi de A. García-Hernandez (2017) [4], utilitza els MFFCs (coeficients cepstrals de freqüència, que modelen com percebem el timbre del so) en el desenvolupament d'un sistema de reconeixement d'activitats humanes.

Cal destacar la importància d'un mètode en el qual es basen la majoria de les mètriques espectrals, la Transformada de Fourier de Curt Temps (STFT), permet obtenir una representació (temps-freqüència) del senyal, essencial per analitzar canvis espectrals al llarg del temps.

L'aprenentatge profund també està present en l'extracció de característiques d'àudio, a diferència de les mètriques anteriors utilitzades per captar propietats concretes del senyal, els models són entrenats per aprendre representacions inspirades en la similitud perceptiva.

OpenL3 es basa en dos articles [5][6], es presenta com una implementació oberta i optimitzada per obtenir embeddings d'àudio en diferents contextos. El model és multimodal no supervisat, format per una xarxa neuronal convolucional profunda que rep com a entrada un espectrograma Mel. El sistema s'entrena per predir si un fragment d'àudio coincideix amb una imatge, el que ajuda a representar sons d'una manera que conserva significat semàntic, no només característiques com el to o la intensitat.

### 3.2 Detecció de patrons

Per agrupar elements similars existeixen diferents mètodes, segons la disponibilitat d'etiquetes a les dades poden ser supervisats i no supervisats. En el cas d'agrupacions supervisades s'usen models per a la classificació o regressió, com SVM, NN... Però en el cas dels no supervisats, la millor opció són els algorismes clustering, que descobreixen estructures o grups latents sense informació prèvia.

L'algoritme K-Means [7] es troba entre els més destacats, s'especifica el nombre de grups  $k$ , assigna a cada punt del conjunt al grup més proper segons la distància al centroid.

El DBSCAN (Density-Based Spatial Clustering of Applications with Noise) [8] es basa en la densitat de punts per formar agrupacions, no és necessari prèviament definir el nombre de clústers i identifica el soroll del conjunt de dades.

El HDBSCAN (Hierarchical DBSCAN) [9] és una extensió del DBSCAN que crea una jerarquia de densitats i selecciona automàticament el nombre òptim de clústers.

### 3.3 Sistemes de recomanació musical

L'àmbit musical és un escenari interessant per implementar els sistemes de recomanació. Entre els tipus de sistemes de recomanació més habituals es troben el filtratge col·laboratiu, que fa recomanacions basades en els com-

portaments d'usuaris similars, el filtratge basat en contingut, que utilitza característiques de productes per trobar similituds, i els sistemes híbrids, que combinen els dos mètodes anteriors per obtenir millors resultats.

Actualment, les aplicacions com Spotify, Apple Music o Youtube fan servir sistemes de recomanació bastant sofisticats per suggerir contingut musical. Encara que fan servir els sistemes híbrids, les recomanacions no són completament precises pel que fa a la similitud sonora entre cançons, sovint són semblants en gènere musical, artista o popularitat. Les llistes de reproducció que ofereix Spotify [10] estan influenciades per les empreses, cosa que afecta directament els resultats de la recomanació.

## 4 METODOLOGIA

En aquesta secció es detalla el procés d'aquest treball, a la Figura 1 es pot veure un esquema de l'arquitectura d'aquest projecte.

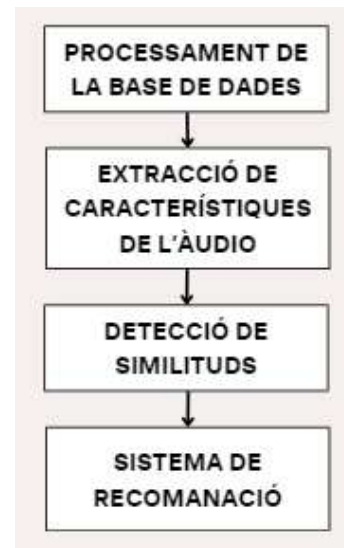


Fig.1: Esquema de la metodologia del treball

### 4.1 Processament de BD

La base de dades està formada per les 200 cançons de la Billboard Global 200 Songs End of Year [11], són les més escoltades l'any 2024. Aquestes cançons han estat processades en fragments de 12 segons, a causa de la diversitat de la cançó, no totes les parts són iguals, i el mètode de detecció de similituds podria no detectar-les.

Per passar-li a la meua companya la informació de les similituds entre les cançons, necessitem convertir els mp3 en arxius MIDI (Musical Instrument Digital Interface)[12]. El format MIDI és un format estàndard per guardar instruccions musicals en format digital. No té àudio, sinó la informació que ens indica com un dispositiu o programari ha d'interpretar una composició musical, alguns aspectes són les notes musicals, la duració de les notes, la intensitat de cada nota, tempo, instruments, entre d'altres. En el projecte de la meua companya, la intel·ligència artificial generativa que implementa utilitza aquest format.

Nom	Valors	Descripció
ID	N/A	Identificador únic de la cançó
song	N/A	Nom original de la cançó
author	188 autors diferents	Nom del(s) intèrpret(s) de la cançó
genres	81 gèneres diferents	Gèneres de la cançó
filename	N/A	Nom de l'arxiu dins la BD
duration	Rang entre [1:27,6:23]	Duració de la cançó
num_splits	Rang entre [7,31]	Nombre de particions de la cançó

TAULA 1. Descripció metadades de les cançons de la BD.

Per recomanar contingut musical similar, no ens basem únicament en el contingut acústic, necessitem poder descriure les cançons originals d'alguna manera. S'ha creat un fitxer de metadades que descriu les 200 cançons triades, a la Taula 1 tenim la descripció estadística de les dades del conjunt.

Genre	Songs	Author	Songs
Pop	55	Billie Eilish	7
Hip-Hop	29	Taylor Swift	6
R&B	16	Ariana Grande	6
Synth Pop	13	The Weeknd	6
Rap	13	Sabrina Carpenter	5
Country Folk	...	Justin Bieber	1
Art Rock	1	Hotel Ugly	1
Pop Rap	1	Glass Animals	1
Rap Rock	1	Frank Ocean	1
Reggae	1	ROSE	1

TAULA 2: Nombre de cançons per gènere musical

TAULA 3: Nombre de cançons per intèrprete

Cada cançó pot tenir diversos artistes, i diversos gèneres, a la Taula 2 es pot observar la distribució de les cançons respecte als gèneres musicals, i a la Taula 3 respecte als autors.

## 4.2 Extracció de característiques

Davant la gran diversitat de mètriques que hi ha per extreure característiques, s'ha implementat tres mètodes formats per diferents mètriques. En aquest apartat es defineixen i s'explica la funció de les mètriques triades.

### 4.2.1 Mètode 1

Per aquest primer mètode s'ha recopilat algunes mètriques que ens aporten informació sobre l'activitat temporal, freqüencial i enèrgic de les cançons. La combinació d'aquestes mètriques ajuda a identificar aspectes musicals com el ritme, la intensitat sonora o el to.

La primera mètrica triada és la velocitat de pas zero (ZCR en anglès), és una eina d'anàlisi en el domini temporal que ens ajuda a calcular la freqüència d'àudio. Mesura la quantitat de vegades que un senyal d'àudio creua l'eix zero, és a dir, canvia de positiu a negatiu.

$$zcr = \frac{1}{T-1} \sum_{t=1}^{T-1} |\text{sgn}[s(t)] - \text{sgn}[s(t-1)]|$$

Fig. 2: Fórmula de la mètrica (ZCR)

Com s'observa a la Figura 2, on es pot veure la fórmula de la mètrica, es comparen els signes de cada parell de mostres consecutives definides com a  $s(t)$  i  $s(t-1)$ , si són diferents significa que el senyal ha creuat el zero, i s'incrementa el valor. Al final es normalitza dividint pel

total de transicions possibles (T-1).

La segona mètrica triada s'anomena centroide espectral, ens indica on se situa el centre de la massa d'un so. En una cançó on tenim freqüències altes al final, el centroide se situarà cap al final, en canvi, en una cançó amb freqüència monòtona, el centroide se situarà en el centre.

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

Fig. 3: Fórmula del centroide espectral

A la Figura 3, es defineix, la fórmula del càlcul del centroide en un instant (t). Per un instant "t" (una columna del espectrograma "S"), es multiplica cada magnitud ( $S[k,t]$ ) per la seva freqüència associada ( $\text{freq}[k]$ ), es sumen tots els productes i es divideix pel total de magnituds en aquell instant ( $S[j,t]$ ).

Com a tercera mètrica s'ha triat l'amplada de banda espectral d'ordre p, és una mètrica de dispersió de l'espectre respecte al centroide espectral. En altres paraules, serveix per determinar el rang de freqüències d'un senyal. Una amplada alta indica un so més complex, en canvi, un nivell més baix, correspon a un so més suau i pur.

$$\text{spectral\_bandwidth}[t] = \left( \sum_k S[k,t] \cdot (\text{freq}[k] - \text{centroid}[t])^p \right)$$

Fig. 4: Fórmula amplada de banda espectral

La banda espectral explicada a la Figura 4 es resumeix en la distància mitjana (elevada a la potència "p") entre cada freqüència ( $\text{freq}[k]$ ) i el centroide espectral ( $\text{centroid}[t]$ ), ponderada per la seva magnitud ( $S[k,t]$ ).

Una altra mètrica triada és l'arrel de la mitjana del quadrat [de l'energia], (RMS), que mesura l'energia de la cançó, és a dir, que tan fort o intensa és en terme mitjà.

$$\text{RMS} = \sqrt{\frac{1}{N} \sum_{i=1}^N x_i^2}$$

Fig. 5: Fórmula del RMS de l'energia

En resum, el càlcul de la fórmula de la Figura 5, consisteix a elevar al quadrat tots els valors del senyal, treure la mitjana i després calcular l'arrel quadrada. El que defineix el volum o potència sonora d'un fragment d'àudio.

Per acabar, es calcula el tempo de la cançó, que es basa en el càlcul de beats per minut.

### 4.2.2 Mètode 2

En aquest mètode es calculen els coeficients cepstrals de freqüència Mel, són un conjunt de característiques (entre 10 i 20 normalment), que descriuen de forma concisa la

forma general de l'espectre basat en la percepció auditiva humana. D'aquesta forma ens aporta informació sobre el timbre de la cançó.

En primer lloc, per calcular els coeficients, se separa el senyal en petits trams. A continuació a cada tram se li aplica la transformada de Fourier discreta per obtenir la potència espectral del senyal. Després se li aplica el banc de filtres corresponents de l'escala Mel [13] a l'espectre del pas anterior, i se sumen les energies de cada espectre.

Finalment, es calcula el logaritme de totes les energies de cada freqüència mel, l'escala del to que percebem els humans, i se li aplica la transformada de cosinus discreta.

Els valors resultants són els coeficients cepstrals de freqüència Mel.

#### 4.2.3 Mètode 3

L'últim mètode d'extracció de característiques es tracta de la funció de la llibreria "*librosa*" [14], *chroma\_stft* aquesta funció calcula un cromatograma a partir d'un senyal acústic. És útil per detectar acords, i identificar la tonalitat de la cançó, mesurant l'energia en les 12 classes de notes musicals independentment de l'octava en la qual es trobin.

Per extreure aquesta informació, es genera un espectrograma dividint el senyal en  $x$  fotogrames. A cada fotograma se li aplica la transformada de Fourier, i utilitzant una escala logarítmica s'associa una nota musical a cada freqüència. Les freqüències que corresponen a cada nota se sumen independentment de l'octava, i obtenim un vector de 12 valors per fotograma de temps, representant la intensitat de cada nota en aquell instant.

### 4.3 Detecció de patrons

Per detectar patrons de similitud entre cançons, s'ha implementat dos mètodes basats en algorismes d'agrupació no supervisats.

#### 4.3.1 DBSCAN

El DBSCAN és un potent algorisme d'agrupació que agrupa els punts que es troben propers dins l'espai de dades. A diferència d'altres algorismes, DBSCAN no requereix l'especificació prèvia del nombre d'agrupacions, cosa que afavoreix l'exploració de dades.

L'algorisme gira al voltant de tres conceptes clau. Els punts centrals que tenen com a mínim un nombre específic de punts (anomenats "*min\_samples*") a una distància determinada (que rep el nom de "*epsilon*"). Els punts fronterers que es troben dins de la distància "*epsilon*" d'un punt central, però no tenen un nombre suficient de veïns "*min\_samples*". Per acabar, els punts de soroll que no es troben prou a prop de cap punt per a ser inclosos.

DBSCAN utilitza dos paràmetres principals, l'*epsilon* que determina la distància màxima entre dos punts per considerar-los veïns. I *min\_samples* que especifica el nombre mínim de punts necessaris per formar una regió densa.

A la Figura 6 situada a l'annex, es detalla el funcionament de l'algorisme amb un diagrama de flux. És important assenyalar que l'elecció dels paràmetres pot afectar significativament els resultats de l'agrupació. I que els paràmetres òptims poden variar segons el conjunt de

dades utilitzat.

Aquest algorisme encaixa molt bé amb l'objectiu que es vol aconseguir, una exploració de la relació que tenen les cançons entre elles, sense especificar prèviament factors determinants com el nombre d'agrupacions.

#### 4.3.2 HDBSCAN

Cal fer l'explicació, els paràmetres i la figura explicativa que anirà a l'annex i serà Figura 7

### 4.4 Sistema de recomanació

Cal buscar informació de com fer el sistema de recomanació, i quins experiments dur a terme per provar el seu funcionament, i amb quines dades.

## 5 DISSENY EXPERIMENTAL

Es plantegen experiments per la comparació dels mètodes de detecció de patrons, per la validació de les agrupacions, i pel funcionament dels mètodes d'extracció de característiques.

### 5.1. Funcionament DBSCAN

El primer experiment busca estudiar el comportament de l'algorisme DBSCAN utilitzant els diferents mètodes d'extracció de característiques i provant diverses combinacions de paràmetres.

L'experiment es basa en una fase exploratòria, on es provaran diferents valors de (*epsilon*) i (*min\_samples*), per observar el comportament de l'algorisme. Per fer l'anàlisi es representaran sis mapes de calor, dos per cada mètode d'extracció. Tres mapes representaran el nombre de clústers generats per cada combinació de paràmetres. I els tres restants representaran el % de soroll per les diferents combinacions de paràmetres.

### 5.2 Funcionament HDBSCAN

S'estudiarà el seu funcionament amb un experiment semblant al realitzat amb DBSCAN.

### 5.3 Validació d'agrupacions DBSCAN

En aquest experiment, es busca detectar quines agrupacions formades per l'algorisme DBSCAN tenen coherència, és a dir, es vol demostrar que el mètode pot detectar patrons similars entre els fragments de les cançons. La hipòtesi d'aquest experiment és que si dues cançons tenen una estructura o sonoritat similar, aleshores alguns dels seus fragments s'agrupen junts dins el mateix clúster.

Per demostrar que l'algorisme funciona, i que els clústers tenen coherència, cada agrupació hauria de contenir principalment fragments d'una mateixa cançó, per la òbvia similitud sonora que presenten, però també pot contenir algun fragment d'una altra cançó que presenta una forta semblança sonora, el que demostraria la hipòtesi plantejada.

Per fer-ho, s'avaluaran els clústers generats pels diferents mètodes d'extracció de característiques, amb dues mètriques de qualitat [15].

La primera mètrica triada és la Puresa dels clústers, ens indica que tan pura és l'agrupació, és a dir, quin percentatge de clúster pertany a una sola cançó.

$$\text{Purity} = \frac{1}{N} \sum_k \max_j |C_k \cap L_j|$$

Fig.8: Fòrmula de la puresa de l'agrupació

La Puresa, definida a la Figura 8, es calcula sumant el nombre de fragments de la cançó predominant ( $L_j$ ) de cada clúster ( $C_k$ ), i es divideix pel total de fragments agrupats ( $N$ ).

La segona mètrica és l'homogeneïtat [16], aquesta mètrica és molt semblant a la puresa, però es calculen de forma diferent, en el cas de la puresa, s'estudia si cada clúster conté majoritàriament una cançó, en canvi, l'homogeneïtat mesura que tan uniformes són els clústers internament, és a dir, si tots els fragments pertanyen a un mateix clúster.

$$h = \begin{cases} 1 & \text{if } H(C, K) = 0 \\ 1 - \frac{H(C|K)}{H(C)} & \text{else} \end{cases} \quad (1)$$

Fig.9: Fòrmula de la homogeneïtat

La fórmula de l'homogeneïtat com s'observa a la Figura 9, es basa en l'entropia, que mesura la incertesa que hi ha entre les classes dins un mateix clúster. Si no hi ha incertesa entre cançons, l'homogeneïtat és 1.

A partir dels resultats de les mètriques, s'exploraran els clústers que tinguin valors més baixos que 1, per analitzar si els fragments agrupats de diferents cançons són semblants i validar la hipòtesi.

Per aquesta part s'avaluaran cada conjunt de cançons semblants dels tres mètodes, utilitzant la mètrica d'alineació temporal dinàmica, *Dynamic Time Warping* (DTW) [17], que permet comparar seqüències temporals.

De cada cançó s'extraurà pel primer mètode el seu espectrograma, pel segon mètode els MFCC, i pel tercer mètode el cromatograma. D'aquesta manera el DTW pot determinar si són semblants o no.

### 5.3 Validació d'agrupacions HDBSCAN

Es farà el mateix experiment realitzat per DBSCAN per validar els clústers.

## 6 RESULTATS

En aquesta secció, s'analitzaran els resultats dels experiments plantejats en el disseny experimental.

### 6.1 Funcionament DBSCAN

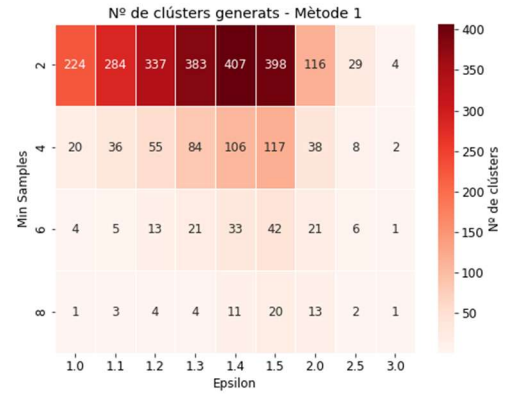


Fig.10: Mapa de calor del nombre de clústers pel primer mètode

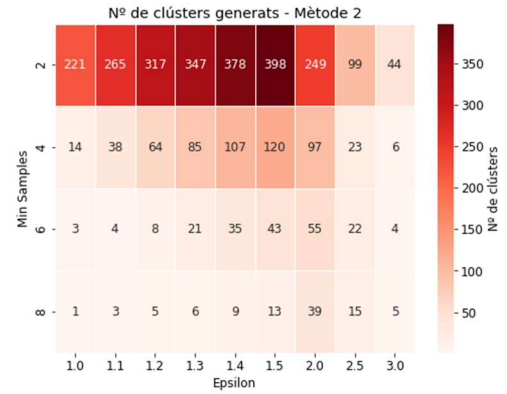


Fig.11: Mapa de calor del nombre de clústers pel segon mètode

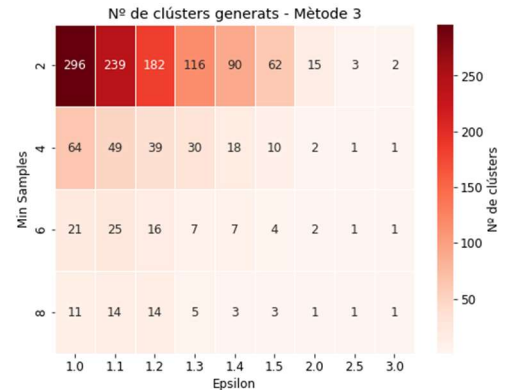


Fig.12: Mapa de calor del nombre de clúster pel tercer mètode

Als mapes de calor de les Figures 10, 11 i 12 es representen la quantitat de clústers formats per l'algoritme DBSCAN amb diferents combinacions de paràmetres.

El primer mètode Figura 10, genera una gran quantitat de clústers, especialment quan el paràmetre *min\_samples* és baix, i l'*epsilon* es troba entre [1.3,1.5]. Es pot veure fàcilment que a partir d'una *epsilon* > 1.5, el nombre de clústers comença a disminuir, el que pot indicar un augment d'agrupació entre els punts.

El segon mètode Figura 11, té un comportament semblant al primer, però presenta una major quantitat de clústers quan l'*epsilon* > 1.5, en comparació del primer mètode.

En últim lloc, el tercer mètode Figura 12 mostra uns resultats diferents dels altres dos. S'observa una menor



generació de clústers en totes les configuracions. El que pot indicar poca variabilitat entre les característiques extretes.

En general, els clústers van disminuint quan augmentem els paràmetres, en el cas d'èpsilon > 1.5, i en el cas de min\_samples > 2. Per completar l'anàlisi analitzarem els resultats de la presència de soroll en les execucions.

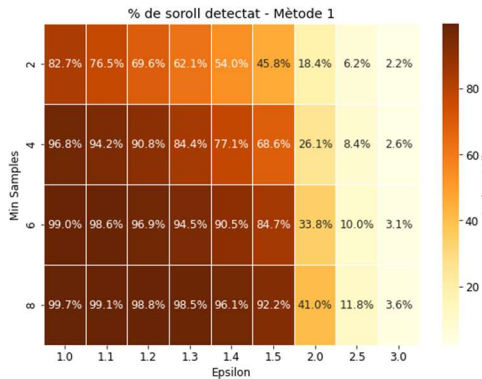


Fig. 13: Mapa de calor de la presència de soroll pel mètode 1

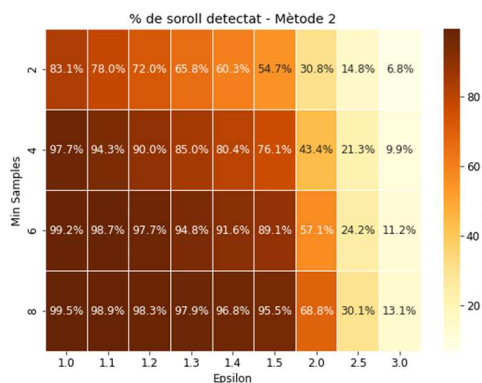


Fig. 14: Mapa de calor de la presència de soroll pel segon mètode.

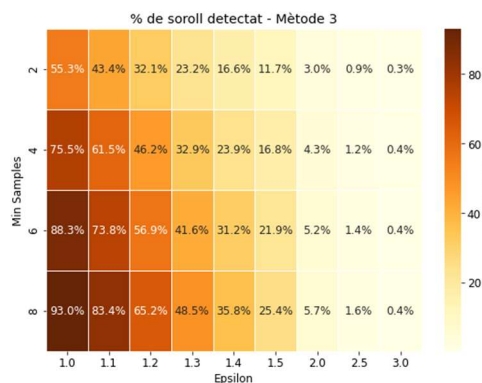


Fig. 15: Mapa de calor per la presència de soroll pel mètode 3.

Als mapes de calor de les Figures 13, 14 i 15 es representa la presència de soroll en les execucions de l'algoritme DBSCAN amb diferents paràmetres.

En el primer mètode Figura 13, podem observar bastant soroll quan l'èpsilon és petit, a partir del valor 1.5 comença a disminuir, al contrari del paràmetre min\_samples que quan creix augmenta soroll.

El segon mètode Figura 14 es pot observar un comportament similar al del mètode 1, per valors d'èpsilon entre [1.0,1.5] i min\_samples entre [2,6] el soroll és major en el segon mètode. Però en valors més grans el soroll és major en el primer mètode.

El tercer mètode Figura 15, com s'ha vist en els mapes de calor del nombre de clústers presenta un comportament diferent dels altres dos mètodes. Es pot observar una menor presència de soroll, el que confirmaria que les dades tenen poca variabilitat entre elles, és a dir, que es poden relacionar fàcilment.

En conclusió, comparant els tres mètodes amb les sis gràfiques representades, es determinen diferents comportaments. Quan l'èpsilon creix en qualsevol dels tres mètodes, es formen menys grups, però hi ha menys soroll, per tant, es formen grups més nombrosos. Quan el min\_samples creix, és a dir, es necessiten més veïns per formar un clúster, es formen menys grups però hi ha més soroll, cosa que determina que no hi ha prou valors similars amb èpsilons petits per generar els clústers.

Per assolir l'objectiu de determinar similituds entre cançons, es necessita una quantitat elevada de clústers, per tant, l'èpsilon no pot ser major a 1.5, en el cas del paràmetre min\_samples, tampoc pot ser molt gran, ha de tenir un valor dins el rang [2,6].

## 6.2. Funcionament HDBSCAN

En aquest apartat explicarem els resultats de l'experiment que busca estudiar el comportament de l'algoritme HDBSCAN de forma similar a l'experiment 5.1

## 6.3. Validació d'agrupacions DBSCAN

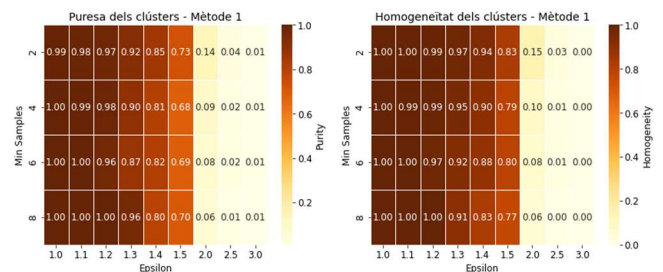


Fig. 16: Mapes de calor per les mètriques de coherència dels clústers.

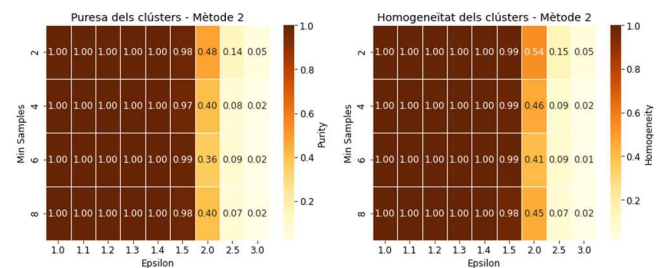


Fig. 17: Mapes de calor per les mètriques de coherència dels clústers.

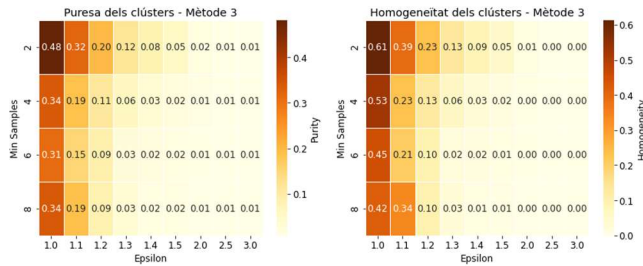


Fig.18: Mapes de calor per les mètriques de coherència dels clústers

A les Figures 16, 17 i 18 estan representades les mètriques de puresa i homogeneïtat que ens indiquen si hi ha una predominança d'una cançó en cada clúster format, el que indicaria coherència en la generació d'agrupacions.

Es pot observar que la diferència entre les dues mètriques és mínima. Pel mètode 1, veiem una homogeneïtat i puresa gairebé màxima en les agrupacions formades amb valors epsilon menors a 1.5.

Pel mètode 2 augmenta el resultat de les mètriques a una puresa i homogeneïtat màxima per qualsevol *min\_samples* amb una epsilon menor a 1.5.

Per acabar, com s'ha analitzat en experiments anteriors, el comportament del mètode 3 és diferent dels altres dos, en aquest cas, el mètode 3 presenta uns valors molt inferiors d'homogeneïtat i puresa, fet que fa dubtar si fiar-se dels resultats de l'algoritme amb aquestes característiques.

En resum, els mètodes 1 i 2 presenten bons resultats d'homogeneïtat i puresa, per tant, es pot dir que detecten correctament fragments musicals similars. El següent pas és analitzar els fragments infiltrats en clústers on predomina una altra cançó, i calcular la distància entre elles. Analitzarem els clústers que han presentat una alta homogeneïtat i puresa.

Davant la gran quantitat de semblances entre fragments diferents, per ara s'avaluarà la semblança de les cançons diferents del mètode 2 amb la combinació de paràmetres *epsilon* (1.5) i *min\_samples*(2), que presenta una puresa del 0.98 i una homogeneïtat del 0.99.

Cançó Dominant (NºFragments)	Altres Cançons (NºFragments)
Song_52 (3)	Song_48 (1)
Song_195 (5)	Song_151 (2)
Song_185 (11)	Song_66 (5)
Song_162 (2)	Song_77 (1)
Song_173 (13)	Song_160 (5), Song_42 (6)
Song_123 (8)	Song_25 (7)

TAULA 4: Cançons semblants del clúster amb el mètode 1

A la Taula 4 es poden observar el nombre de fragments similars de cançons diferents que hi ha als clústers agrupats per l'algoritme DBSCAN configurat amb els paràmetres *epsilon* = 1.5 i *min\_samples* = 2, amb les característiques del mètode 2.

Per saber si els resultats de la mètrica DTW són baixos o no, es calcularà la mètrica DTW entre fragments de la mateixa cançó per tenir un llindar de referència.

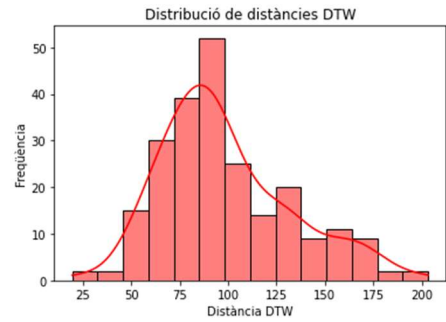


Fig.19: Histograma de distàncies DTW entre fragments similars de la mateixa cançó

A la Figura 19 es pot observar l'histograma de distàncies DTW entre fragments similars de la mateixa cançó, els valors van des de 25 a 200, aquest serà el llindar per determinar si els fragments detectats com a similars de cançons diferents tenen semblança o no.

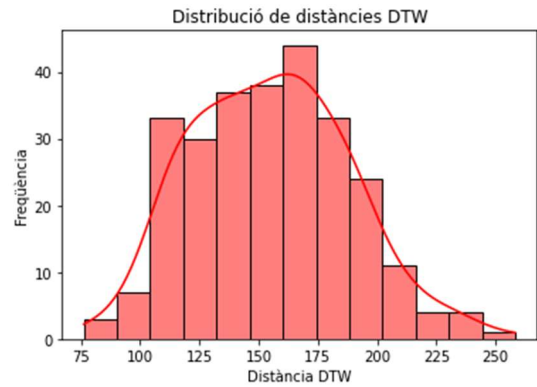


Fig.20: Histograma de les distàncies DTW entre fragments similars de diferents cançons.

A la Figura 20 es mostra l'altre histograma de distàncies DTW, però aquest representa les distàncies entre fragments detectats com a similars de diferents cançons. Els valors van de 75 a 250, i es veu una clara freqüència predominant entre el 150 i el 175.

Comparant els dos histogrames, les distàncies entre fragments de la mateixa cançó són més baixes que en cançons diferents, però no hi ha una diferència molt gran, inclús hi ha fragments que tenen una distància baixa com els fragments de la mateixa cançó.

I això permet afirmar la hipòtesi plantejada, on s'afirma que hi ha fragments de cançons diferents que l'algoritme DBSCAN ha agrupat que tenen una relació acústica evident, i per tant hi ha un patró entre cançons d'èxit.

## 7 CONCLUSIÓ

Per concloure el segon seguiment, en aquesta secció s'explicarà que s'ha aconseguit fins ara, que s'ha plantejat aconseguir en els pròxims dies, i quins dubtes han sorgit en el desenvolupament d'aquest treball.

Seguint les fases establertes en la planificació i la metodologia, s'ha aconseguit crear i processar la base de dades, per poder desenvolupar el projecte.

S'ha investigat diferents mètodes d'extracció de carac-

terístiques acústiques, dels quals se n'han implementat tres.

A més, s'ha estudiat com detectar patrons de similitud entre cançons, fent servir un algoritme capaç de trobar relació entre els fragments de les cançons. Tant els mètodes d'extracció de característiques com els de detecció de similituds han estat avaluats amb diferents mètriques, amb l'objectiu d'observar el seu comportament, i obtenir resultats favorables.

Amb això, es pot considerar que s'ha assolit el primer objectiu del treball: detectar i analitzar patrons sonors entre cançons a partir de les seves característiques acústiques i estructurals, explorant diferents mètodes de processament de senyal, i aprenentatge computacional.

Pel que fa a les tasques previstes per als pròxims dies, vull implementar un altre mètode d'agrupació, concretament una extensió del DBSCAN (HDBSCAN) per analitzar el seu comportament i observar quines relacions detecta entre les cançons.

A més, vull implementar un altre mètode d'extracció de característiques basat en aprenentatge profund, el qual he explicat a l'estat de l'art, concretament OpenL3. Tot i que no he pogut implementar-ho, a causa del seu elevat temps d'execució.

Finalment, pel que fa als dubtes i problemes actuals, principalment destaca la dificultat d'implementar i avaluar un sistema de recomanació, puc crear un sistema de recomanació, però no tinc dades per provar el seu funcionament. A més, també tinc dificultats per saber com interpretar i explicar els resultats de similituds entre cançons: tot i que s'ha detectat una similitud entre fragments, cal trobar una manera clara de descriure aquesta similitud.

## BIBLIOGRAFIA

- [1] J. Hernández. "Cómo el algoritmo de Spotify influye en el consumo de música - Industria Musical". Industria Musical - Otro sitio realizado con WordPress. Accedido el 20 de abril de 2025. [En línea]. Disponible: <https://industriamusical.com/como-el-algoritmo-de-spotify-influye-en-el-consumo-de-musica/>
- [2] K. Mahajan. "Extracting audio features using Librosa". Medium. Accedido el 20 de abril de 2025. [En línea]. Disponible: <https://kaavyamaha12.medium.com/extracting-audio-features-using-librosa-3be4ff1fe57f>
- [3] M. Fuentes Méndez and J. M. Gutiérrez Arriola, "Reconocedor automático de patrones musicales," Proyecto Fin de Carrera/Grado, E.T.S.I. y Sistemas de Telecomunicación (UPM), 2020.
- [4] A. García-Hernandez et al, "A Similarity Analysis of Audio Signal to Develop a Human Activity Recognition Using Similarity Networks," Sensors, vol. 17, (11), pp. 2688, 2017. Available: <https://www.proquest.com/scholarly-journals/similarity-analysis-audio-signal-develop-human/docview/1977885562/se-2>. DOI: <https://doi.org/10.3390/s17112688>.
- [5] Look, Listen and Learn More: Design Choices for Deep Audio Embeddings Aurora Cramer, Ho-Hsiang Wu, Justin Salamon, and Juan Pablo Bello. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP), pages 3852-3856, Brighton, UK, May 2019.
- [6] Look, Listen and Learn Relja Arandjelović and Andrew Zisserman IEEE International Conference on Computer Vision (ICCV), Venice, Italy, Oct. 2017
- [7] K. Babitz. "Introduction to k-Means Clustering with scikit-learn in Python". DataCamp.com. [En línea]. Disponible: <https://www.datacamp.com/tutorial/k-means-clustering-python>
- [8] R. Kumar. "Guía del algoritmo de agrupación DBSCAN". DataCamp.com. [En línea]. Disponible: <https://www.datacamp.com/es/tutorial/dbscan-clustering-algorithm>
- [9] "HDBSCAN". scikit-learn. Accedido el 20 de abril de 2025. [En línea]. Disponible: <https://scikit-learn.org/stable/modules/generated/sklearn.cluster.HDBSCAN.html>
- [10] "Seguridad y Centro de Privacidad". Como funcionan las recomendaciones.[En línea]. Disponible: <https://www.spotify.com/es/safetyandprivacy/understanding-recommendations>
- [11] "Billboard Global 200 Songs". Billboard. Accedido el 20 de abril de 2025. [En línea]. Disponible: <https://www.billboard.com/charts/year-end/billboard-global-200/>
- [12] Colaboradores de los proyectos Wikimedia. "MIDI - Wikipedia, la enciclopedia libre". Wikipedia, la enciclopedia libre. Accedido el 20 de abril de 2025. [En línea]. Disponible: <https://es.wikipedia.org/wiki/MIDI>
- [13] Contributors to Wikimedia projects. "Escala Mel - Viquipèdia, l'enciclopèdia lliure". Viquipèdia. Accedido el 20 de abril de 2025. [En línea]. Disponible: [https://ca.wikipedia.org/wiki/Escala\\_Mel](https://ca.wikipedia.org/wiki/Escala_Mel)
- [14] "Feature extraction — librosa 0.11.0 documentation". Librosa. Accedido el 20 de abril de 2025. [En línea]. Disponible: <https://librosa.org/doc/main/feature.html>
- [15] "Evaluation Metrics for Clustering Models | Towards Data Science". Towards Data Science. Accedido el 20 de abril de 2025. [En línea]. Disponible: <https://towardsdatascience.com/evaluation-metrics-for-clustering-models-5dde821dd6cd/>
- [16] "homogeneity\_score". scikit-learn. Accedido el 20 de abril de 2025. [En línea]. Disponible: [https://scikit-learn.org/stable/modules/generated/sklearn.metrics.homogeneity\\_score.html](https://scikit-learn.org/stable/modules/generated/sklearn.metrics.homogeneity_score.html)
- [17] Contributors to Wikimedia projects. "Dynamic time warping - Wikipedia". Wikipedia, the free encyclopedia. Accedido el 20 de abril de 2025. [En línea]. Disponible: [https://en.wikipedia.org/wiki/Dynamic\\_time\\_warping](https://en.wikipedia.org/wiki/Dynamic_time_warping)



APÈNDIX

A1. FIGURES EXPLICATIVES

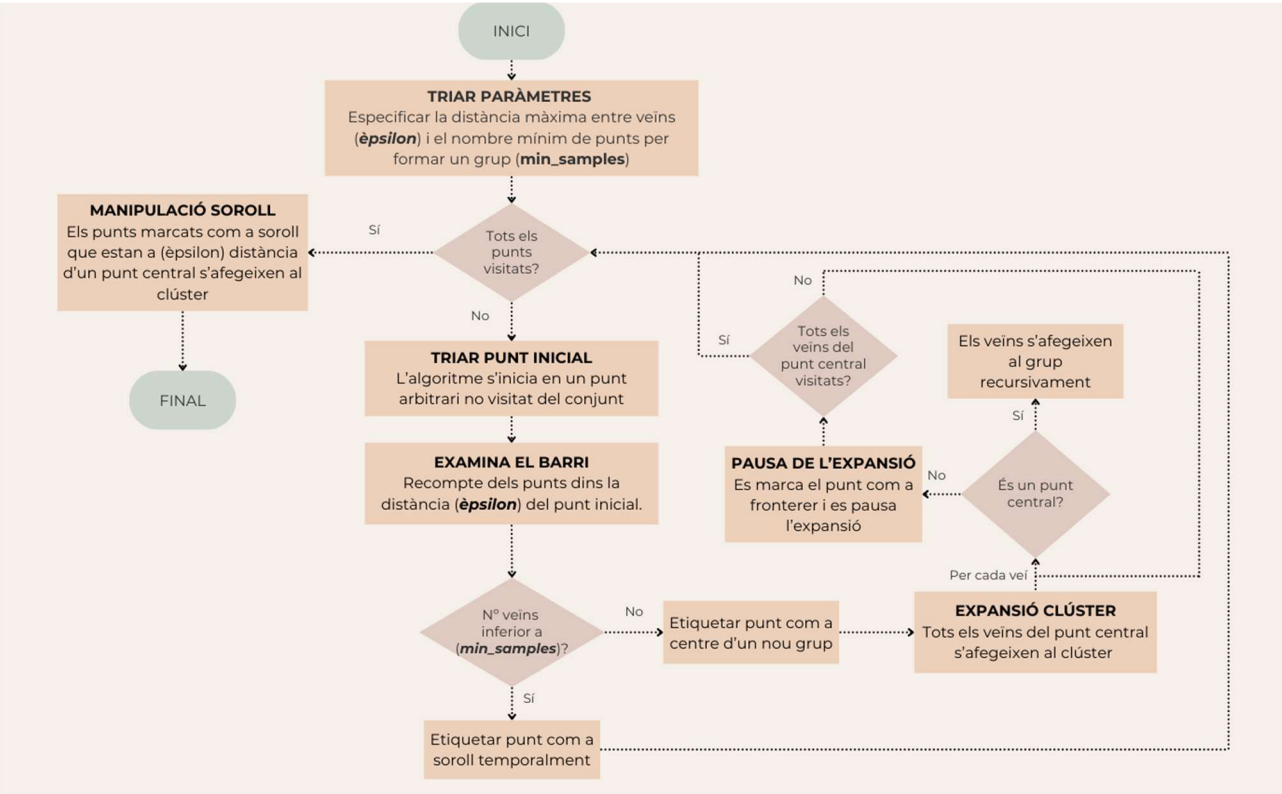


Fig.6: Diagrama explicatiu de l'algoritme DBSCAN