Classificació de gèneres musicals a partir de Machine Learning

David Piñol Navas
Barcelona, Spain

Abstract

La capacitat que té la música per afectar a les emocions la fa àmpliament reconeguda com a un llenguatge universal. El concepte de "gènere musical" es fa servir amb freqüència per agrupar diversos estils musicals que segueixen unes mateixes pautes. L'objectiu d'aquest treball és el de categoritzar eficientment diferents gèneres musicals a través de tècniques de machine learning.

Music Genre Classification, Audio Features, Machine learning

0.1 Introducció

Vivim en una època en la qual l'enorme nombre de cançons que podem trobar en les plataformes d'estreaming és major que mai, això significa que el número de gèneres musicals està també augmentant al mateix ritme. Per tant, pot generar bastant caos per alguns consumidors, intentar organitzar tota aquesta música i distingir un número tan gran de cançons de manera manual. Per aquesta raó, un mètode per poder-les classificar automàticament pot ajudar als usuaris a poder categoritzar millor la seva música i també poder entendre millor les relacions intrínseques entres els diversos gèneres.

0.1.1 Elecció del Dataset

Per acomplir aquesta tasca, farem ús del Dataset GTZAN, el qual podem trobar a la pàgina de Kaggle: `https://www.kaggle.com/datasets/andradaolteanu/gtzan-dataset-music-genre-classification?datasetId=568973&sortBy=voteCount`

El dataset està format per 10 gèneres, els quals consta de 100 àudios de 30 segons, de 16 bits, 22050Hz i en mono. Conté els gèneres ja preclassificats els quals són: Blues, Jazz, Metal, Pop, Reggae, Disco, Classical, Hip-hop, Rock and Country.

0.1.2 Definició de variables

- filename: nom del fitxer.
- length: duració de l'àudio (ms).
- chromaStft: canvis tonals basats en la transformada de Fourier. Indica quines són les freqüències més repetides al llarg de tota la mostra i, per tant, quines són les notes que més vegades sonen, això ens serveix per poder veure si és una cançó molt o poc tonal.
- rms: root mean square, indica el volum real de la cançó.

- spectralCentroid: indica el centre de masses de l'espectrograma, és a dir, quina és la freqüència "mitjana". Com més alt sigui el valor, més agut i brillant serà la cançó i com més baix, més greu i fosc.
- spectralBandwidth: diferència entre les freqüències superior i inferior en una banda contínua de freqüències.
- rollof: indica els límits de l'espectre.
- zeroCrossingRate: és la freqüència amb la qual el senyal passa de positiu a negatiu o viceversa. És un indicador útil, ja que ens pot estar informant de la quantitat de soroll que hi ha.
- harmony: indica l'energia dels harmònics en funció del temps i la freqüència.
- perceptr: ponderació perceptiva rítmica.
- tempo: bpm de la cançó.
- mfcc1-mfcc20: Coeficients Cepstrals en les Freqüències de Mel. Són coeficients en diferents trams de l'espectre a partir de la transformada de Fourier i un filtre perquè estigui basat en la percepció auditiva humana.

0.2 State of art

Existeixen com a mínim 100 autors (només mirant Kaggle) que han treballat amb aquest dataset i han entrenat un model predictiu, nosaltres, però, només ens fixarem en els tres autors més rellevants [1]. Si veiem els seus notebooks, veiem que han escalat les dades amb MinMax Scaler i Standard Scaler. Cadascú ha obtingut el millor resultat amb diferents models, Convulotional Neural Networks (0.92), XBoost (0.90) i Random Forest(0.81). Cal mencionar que cap d'aquests autors ha fet cap neteja de dades ni un preprocessat concret a part de l'escalat.

Com podem veure en [2] es fa una comparativa entre l'efectivitat d'utilitzar els coeficients basats en Fast Fourier Transform i els de Mel Frequeny Ceptrum, el qual no fa una anàlisi lineal de les freqüències sinó logarítmic, de forma que intenta imitar la forma en la qual l'oïda humana percep el so. És

per això que el dataset GTZAN proporciona aquests coeficients.

Existeixen algunes eines interessant que podem trobar a internet que intenten donar una solució al problema que estem mencionant.

EveryNoise [3], es tracta d'una pàgina web que ens mostra més de 5000 gèneres musicals ordenats de més electrònics a més orgànics i que presenta un diagrama de dispersió en el que permet entrar a cada gènere i escoltar un fragment d'una cançó que correspon al gènere seleccionat; veure els artistes més representatius de cada un, i poder buscar un artista i veure els diferents gèneres musicals que el defineixen.

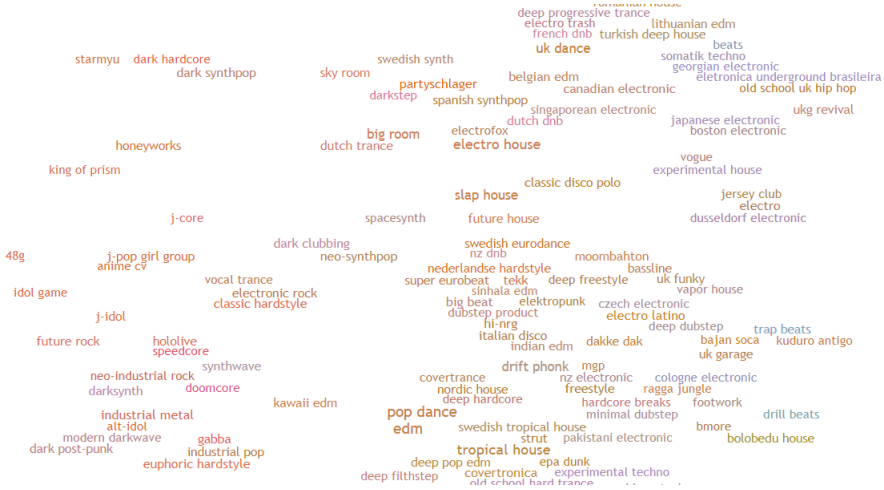

Figure 1:

Chosic [4], ofereix diferents eines vinculades a la música. com per exemple, dir-te si una cançó és molt alegre, trista, ballable, etc. L'eina que més ens interessa és l'anomenada 'Music Genre Finder' Un buscador que promet dir-te el gènere d'una cançó, artista o llista de reproducció. És una funcionalitat que no va malament, però únicament recull informació de Spotify i Wikipedia per saber el gènere en qüestió, en lloc d'utilitzar tècniques de ML que serien més precises.

getGenre() [5], és una pàgina molt més simple que les anterior. Té un petit buscador en el qual a l'introduïr un artista, mostra tots aquells gèneres als quals correspon. La pàgina assegura que funciona gràcies a una precisa Intel·ligència Artificial, però la nostra experiència és que és molt genèric i gens precís.

0.3 Visualització de les dades

En la Figura 1 podem veure la relació de cada gènere amb la variable de chroma_stft_mean, la qual, com hem dit abans, explica si hi ha moltes notes que es repeteixen sovint, aquest concepte està estretament relacionat amb el centre tonal d'una cançó, quan una melodia té una certa tendència a repetir una mateixa nota direm que el seu centre tonal està en aquella nota. Doncs bé com més evident sigui aquest centre tonal més alt serà el valor de chroma_stft, ja que unes determinades notes s'estan reproduint més vegades que altres, mentre que si el valor de chroma_stft és baix, significa que no hi ha un centre tonal clar, perquè, per exemple, l'harmonia modula amb freqüència a diferents tonalitats i, per tant, no permet que hi hagi una melodia amb un centre tonal definit. Cal mencionar que el 'pitch' d'una nota es pot separar en dues coses: l'altura tonal i el chroma, per això, el chroma no farà distinció de l'octava de la nota.

Si tornem a mirar la Figura 1, veiem que els dos gèneres amb el valor més baix són la música clàssica i el jazz, cosa que té sentit, ja que tenen més modulacions tonals i, per tant, els que tenen una harmonia més complexa.

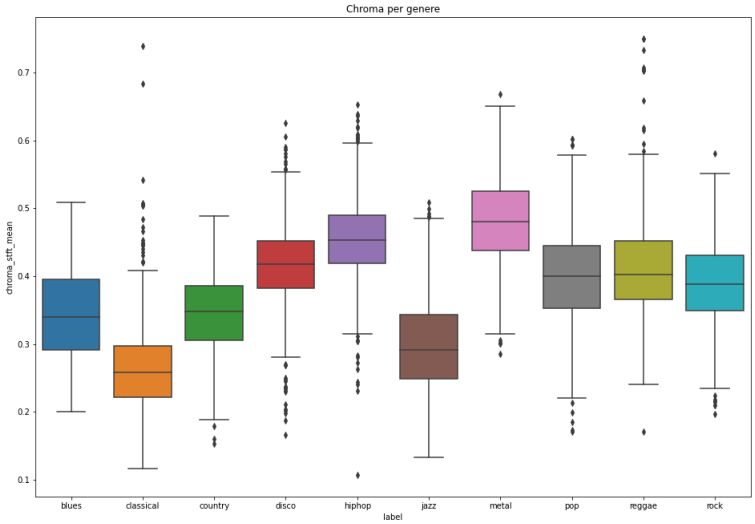

Figure 2: Boxplot de chroma_stft de cada gènere

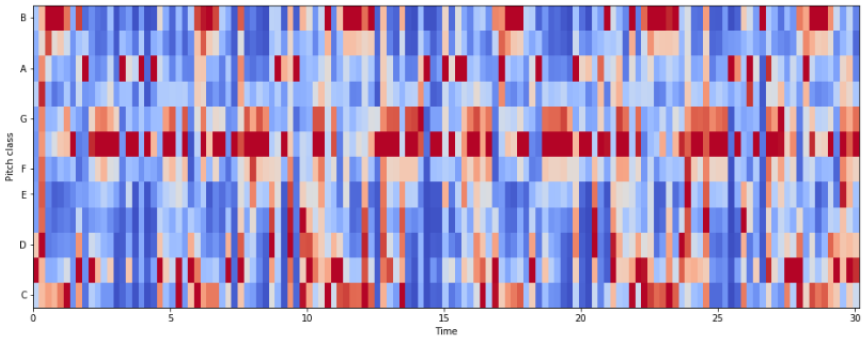

Figure 3: Visualització del chroma d'una mostra

0.4 Metodología

0.4.1 Preprocessat

No trobem dades nul·les ni repetides, així que no caldrà eliminar ni omplir cap dada. Variables com filename i lenght seran eliminades, ja que no es poden utilitzar per entrenar el model

Si mirem la correlació de la base de dades, podem veure que hi ha més d'un 95% de correlació amb el spectral centroid i el spectral bandwidth respecte del rollof. De tal manera que descartarem aquestes dues variables, ja que no ens ajudaran a entrenar el model.

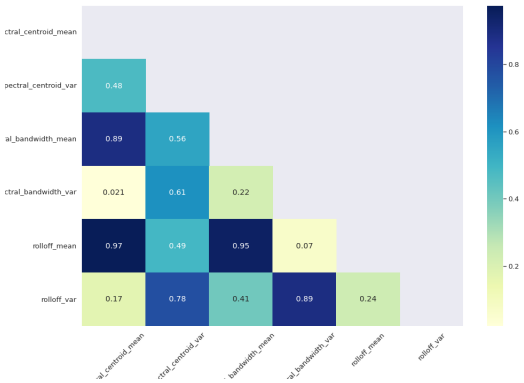

Figure 4:

Tot i ser conscients que tenim ouliers en les variables, hem vist que els millors resultats els hem aconseguit escalant les dades amb MinMax Scaler en lloc de Robust Scaler, per tant, és el que farem servir.

0.4.2 Creació del Model

A l'hora de dividir la base de dades en train i test, hem decidit utilitzar la proporció de 20/80 en lloc de la de 30/70 com podem veure en els treballs mencionats.

Hem decidit realitzar el test amb 9 models diferents per veure quin és el que obté millor puntuació i així poder-lo comparar amb els resultats dels models dels altres autors.

0.5 Resultats

El model que ha obtingut la millor puntuació ha estat el de KNeighbornsClassifier, amb un 90.34% d'accuracy (Fig.5).

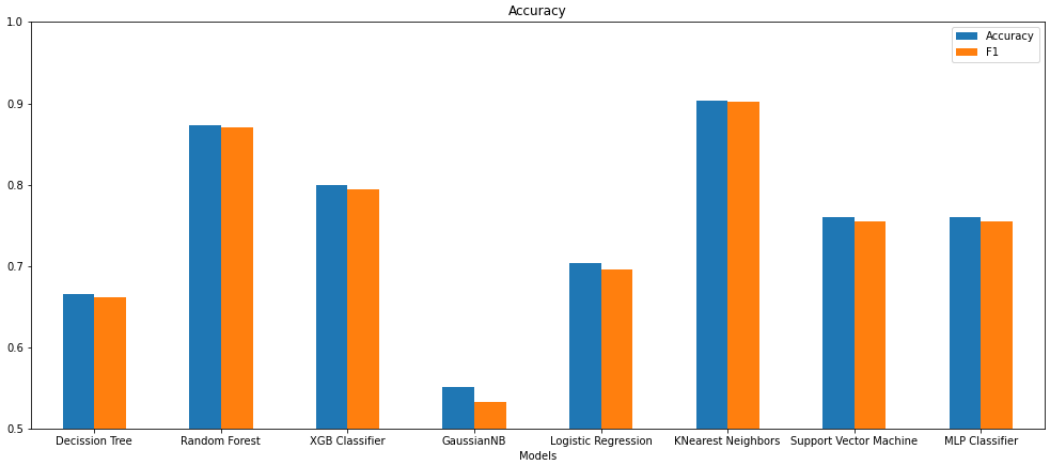

Figure 5:

Arribats a aquest punt ja hem aconseguit una major puntuació en accuracy que el que podem trobar en la majoría de notebooks de Kaggle, però encara podem buscar els millors paràmetres del millor model per tal de millorar-lo. Per fer-ho utilitzarem Random Search i tornarem a entrenar

el model amb aquells paràmetres. Per altra banda, per evitar sobreentrenament aplicarem una validació creuada passant-li per paràmetre el número de 'folds' que volem que ens faci dins de la funció de Random Search. Una vegada ja hem trobat els hyperparàmetres, els hem aplicat i hem fet la validació creuada, podem veure que hem obtingut una puntuació de fins a 93.29% (Aquest resultat por variar lleugerament cada vegada)

Els resultats de la matriu de confusió obtinguts gràcies a al model de K Nearest Neighbors es poden veure en la figura 6.

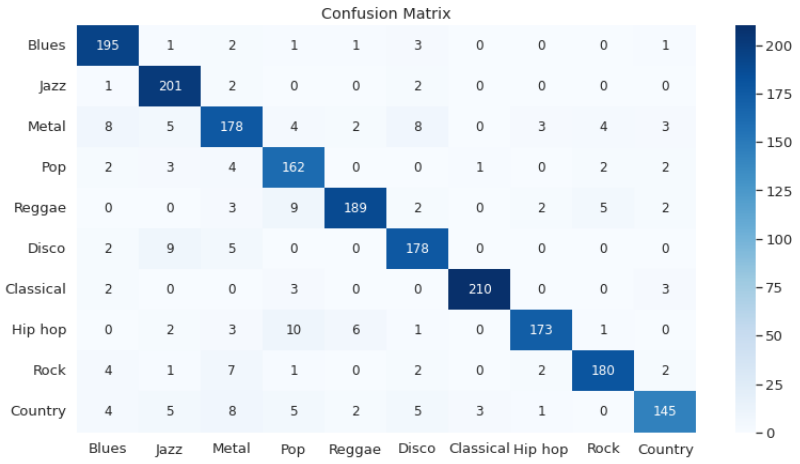

Figure 6: Matriu de confusió

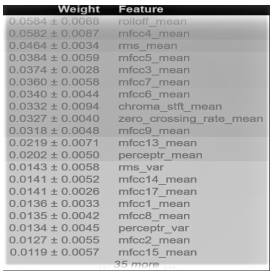

Figure 7: Feature importance

Podem veure en la figura 7 que pel que fa a la rellevancia de les variables,

les tres més rellevants han estat rolloff_mean, mfcc4_mean i rms_mean.

0.6 Conclusió

La classificació ha estat sempre un repte i molts treballs han estat publicats sobre el tema en els últims anys. En aquest treball s'ha intentat comparar els resultats d'altres models de machine learning de Kaggle i s'ha pogut obtenir una major precisió en el model gràcies a l'algoritme de K Nearest Neighbors i Randon Search per obtenir hyperparàmetres.
Com hem vist en l'estat de l'art, no hi ha com a tal una eina en la qual a partir de tècniques de machine learning, pugui dir-te el/els gèneres que defineixen una cançó determinada. És per això que seria interessant que poguéssim entrenar un model amb una base de dades més gran, com la que podem trobar a EveryNoise [3], amb més de 5000 gèneres, per tal que a partir d'una mostra d'àudio, a l'estil Shazam, pugui dir-te a quin gènere correspon.

Bibliography

[1] [Briot, J. P., Hadjeres, G., & Pachet, F. D. (2017). Deep learning techniques for music generation–a survey. arXiv preprint arXiv:1709.01620.]

[2] Batucan Senkal - Data Scientist at Veloxity Inc - Istambul, Istambul, Turkey. https://www.kaggle.com/code/psycon/audio-data-eda-processing-modeling-recommend

Padmavathi - Analyst - Chenna, Tamil Nadu, India. https://www.kaggle.com/code/aishwarya2210/let-s-tune-the-music-with-cnn-xgboost

Andrada Olteanu - Data Scientist at Endava - Bucharest, Bucharest, Romania. https://www.kaggle.com/code/andradaolteanu/work-w-audio-data-visualise-classify-recommend

[3] EveryNoise: https://everynoise.com/engenremap.html

[4] Chosic: https://www.chosic.com/

[5] getGenre(): https://www.getgenre.com/