Studio SMP 🤯

MUSIC ☠

GENRE CLASSIFYING SYSTEM

STUDIO SMP

WOT IS IT

- A music genre classification and education system
- It uses machine learning to analyze audio files, predict their musical genre(keep in mind that machine learning is basically about learning from examples and forming patterns.)
- Provides simple explanations for how and why it reached those conclusions. (Like a human-AI partnership😌)
- Making technical audio analysis accessible to musicians, educators, or learners(also keep in mind that music is just patterns in vibration)

OVERVIEW

1

Trained model (XGBoost) predicts one of four genres: Classical, Jazz, Rock, or Disco.

2

Shows real-time confidence scores for predictions.

3

Currently at 82% test accuracy using a subset of the GTZAN dataset.

4

Confidence scores are shown for all four genres- The final predicted answer is of the one with highest confidence value

KEY FEATURES

1

Extracts 34 features using the librosa library, grouped as:

- Rhythm & Timing (tempo)
- Voice & Instrument Texture (MFCC 0–4)
- Musical Style (MFCC 5–8)
- Energy & Dynamics (MFCC 9–12)
- Musical Notes & Harmony (chroma 0–11, tonnetz)
- Sound Color & Texture (spectral contrast 0–6)

2

- Uses SHAP (SHapley Additive exPlanations) to illustrate which features most influenced each prediction.
- Provides visual feature importance rankings.
- Bridges technical analysis and musical intuition for users.

3

For users looking to understand music technicalities:

- Simplifies audio processing concepts with analogies
- Explanations clarify technical terms for non-specialists
- Results include not just predictions, but why the AI chose them

4

About the UI:

- Web interface for uploading and analyzing audio
- Intuitive feedback
- Supports MP3, WAV, and M4A files (up to 50MB).

(READ IF CURIOUS ABOUT THE UNDERLYING LOGIC)

Part 1

Digitized music comes down just as a stream of numbers(audio samples). Different music genres basically represent differences in these statistical patterns. (for example, in rhythm, tempo, harmonies, instrumental tone,and a lot more.)
By converting audio physics into numbers that reflect music theory concepts, you allow ML to spot the same differences humans do, just in a more abstract form.

Feature extraction is in the simplest sense, a conversion of the physics properties into numbers. Happens through signal processing methods. So the model doesn't hear music, instead, it works on these numbers(imagine properties like:

- how energy patterns are distributed across frequences
- the harmonic relations data
- variation in the sound energy spectrum

All these features that explain the entire audio, are then stored(encoded!)into a vector of numbers for each song.

Labelled training data is what the model uses to get our stuff done- it uses those audio files with "ground truth" genre labels, keeps analyzing each file- what features combinations eventually make up a genre, and given enough time, it can make accurate generalizations.

(READ IF CURIOUS ABOUT THE UNDERLYING LOGIC)

Part 2

Audio file is loaded-->signal processing is done by libraries like librosa-->eventually 34 features extracted- encoded in the statistical summaries we want. -->features are standarized so that any random musical properties are not artficially given uneven importance.
All models follow different algorithms.

Ours is xgboost- a tree based algorithm: it tries various (decision trees) splits to map physical properties to genres, by answering questions that can potentially classify genre properties.
Feature distributions for genres can overlap, that's why the model tries to set useful boundaries.
So by learning on enough labelled data it can finalize on the most relevent questions.

So when it gets a new audio clip, it tries to extract these features and feeds it to the model which tries to compute genre. SHAP analysis helps us see which features were most useful for the classification. All this is based off our underlying assumption that different music pieces do have distinct physically measureable and computable properties.
So this is what a physics informed statistical sorting process looks like.

COOL MUSICIAN VS ML MODEL...?

COOL MUSICIAN

- requires sound+ context +emotion +memory from previous experiences.

- Parallel+non-linear+ integrative+intuitive processing style.
- Can be fooled by unfamiliarity or without solid context.

ML MODEL

- Requires extracted numerical features only-learns mapping from labelled data.
- Systematic splitting of possibilites+ numerical processing style.
- Can be fooled if there are no proper boundaries between musical properties of genres.

True musical innovation= musician's intuition+emotional nuance+ other crazy human brain artistry+ ML's analytical precision and speed in pattern processing. 🤯