Is It Techno or House? Let's End the Debate (With Science!)



One day, you're minding your own business, scrolling through the internet, when suddenly—bam!—you find yourself in a heated argument. Some random stranger insists that the track you're vibing to is

localhost:8501 1/10

definitely Techno, but you know it's House. Or is it? Panic sets in. What if your listening history isn't impressive enough to back up your claim? What if you're about to get ratioed into oblivion?

But wait—before you start second-guessing your entire music taste, let's take a step back. EDM is a vast, wonderfully chaotic world where genres blend, mutate, and sometimes defy definition altogether. Some tracks sit neatly in one category, while others laugh in the face of classification. So, instead of relying on vibes alone, why not fight back with cold, hard data?

Welcome to the ultimate breakdown of what makes House House, Techno Techno, and Dubstep... well, an earthquake with bass. We're diving deep into spectral features, machine learning, and genre clustering to uncover the science behind your favorite beats. By the end of this, not only will you have an arsenal of data-driven arguments at your disposal, but you might even discover why some genres refuse to be put into neat little boxes.

Dataset Source Explanation

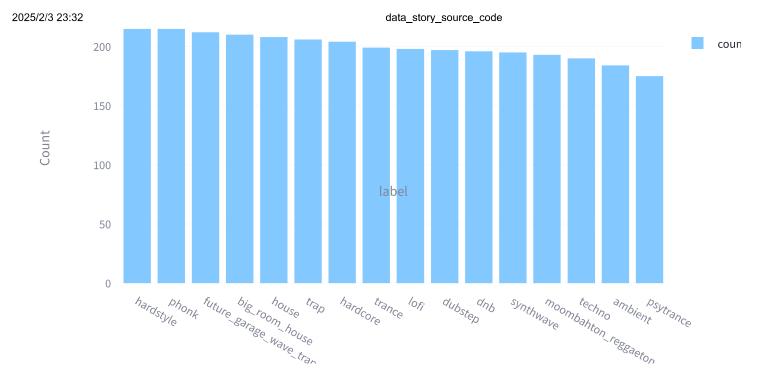
For this project, we are using the EDM Music Genres Dataset from Kaggle, originally uploaded by Sivadithiyan (https://www.kaggle.com/datasets/sivadithiyan/edm-music-genres). This dataset is sourced from YouTube music mixes and contains 16 different electronic music genres (such as Techno, House, Trance, etc.). Each genre is represented by 2,500 three-second audio clips, evenly split into 2,000 for training and 500 for testing. The dataset was processed using Ableton, extracting spectral features, MFCCs, Chroma, Tonnetz, and other audio characteristics for music classification, genre analysis, and machine learning applications.

However, because the full dataset is very large, using it directly might lead to performance issues (or, in extreme cases, turn your CPU into an expensive toaster). To make the program run smoothly on most machines, we use a smaller dataset that randomly selected 10% subset of the training data.

If you have a powerful computer (and don't mind your fans sounding like a jet engine), you can swap out train_sample.csv for the original train_data_final.csv in the code. The program is designed to handle this switch seamlessly.



Genre Count



What are those? To me, these are just a bunch of bars that look almost the same height.

Here, we have a bar chart showdown of EDM genres, where each genre is flexing its musical muscles to prove who dominates the dataset! Think of it like a popularity contest at a massive festival—except instead of fans, we're counting tracks. Looking at the chart, it's clear that no single genre is stealing the spotlight—most genres have roughly the same number of tracks. This means we've got a balanced dataset, so every genre gets a fair shot at being analyzed.

Sounds interesting. But what if some of them is much taller than others?

If one genre had way more tracks than the others, our analysis might be biased, like having a DJ that only plays House music all night (that would be so boring). Since the dataset is evenly spread, our machine learning models can learn fairly from all genres—no favoritism here! If your favorite genre is at the bottom of the chart, don't worry—it's about quality, not quantity.

Well, as an Ambient fan I'm glad to hear this.

What's Next? Now that we know how our dataset is structured, let's dig deeper—what makes these genres sound different?



Spectral Features by Genre

If EDM were a high-tech cooking show, then spectral features would be the ingredients that define the flavor of each genre. Some genres love spicy high frequencies (like Techno), while others prefer smooth and creamy bass (like Lofi). Just like a chef carefully selects ingredients, producers tweak these spectral properties to craft the perfect sonic experience.

localhost:8501 3/10 Each genre has its signature sound. For example:

House & Techno has higher spectral centroid which determines brightness (like how much spice in a dish).

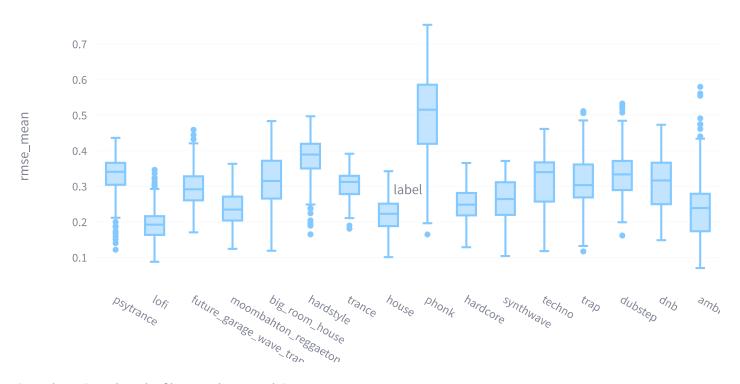
Lofi & Ambient has lower rolloff which indicates energy in low frequencies (like adding extra butter for richness).

Dubstep & Drum & Bass has wider spectral bandwidth which measures spread of frequencies (like how complex the flavors are).

Select a feature to compare:

rmse_mean v

rmse_mean by Genre



I'm only seeing a bunch of boxes, please explain

Think of this box plot as a musical talent show—each genre steps onto the stage, showing off how its spectral feature varies. Some genres are consistent performers, while others have wildly experimental tracks that don't fit the norm.

Here's how to read the stage performance:

localhost:8501 4/10

The box represents the middle 50% of values—this is where most tracks in a genre hang out. The line inside the box is the median—the "average performer" for that genre. The whiskers show the full range of normal values—stretching their arms out to show flexibility. Dots outside the whiskers? These are the rebels—tracks that push genre boundaries and break expectations!

Taking Root Mean Square Energy (RMSE) as an example, it measures the loudness fluctuations within a track. A higher RMSE generally indicates a track with punchy, dynamic energy, while a lower RMSE suggests a softer, more consistent sound. If we click on the bar and take a closer look at the selected spectral feature rmse_mean and see how different genres perform, we can find out that Phonk is like that contestant who can do everything from opera to rap—it has a huge range of values, meaning tracks in this genre vary a lot in spectral characteristics. Lofi & Psytrance are the disciplined ones—their values are tightly packed, meaning these genres have a consistent sound that artists stick to. Dubstep & Ambient have some rule-breakers—those outliers could be experimental tracks or hybrids that mix with other genres.

Due to space limitations, we can't walk you through every selectable feature from the dataset, but feel free to explore them all on your own!

So, what's the conclusion? Do we get anything that the data didn't explicitly tell us?

The most important conclusion is: genres are not mololithic. Even within a single genre, spectral properties vary widely, and some tracks blur the boundaries between styles. If you were a producer and you want to make your track "more techno", you can tweak spectral centriod, MFCCs and RMSE to match the genre's signature.

Now that we've learned individual features, what about putting them together? Can we cluster genres based on their spectral properties? Let's move on to find the answer.



Clustering of EDM Genres

To understand the relationships between EDM genres, we need a way to visualize high-dimensional data in a meaningful way. Songs have many spectral features, but humans can only intuitively interpret 2D or 3D graphs. This is where dimensionality reduction techniques like PCA, UMAP, and t-SNE come in.

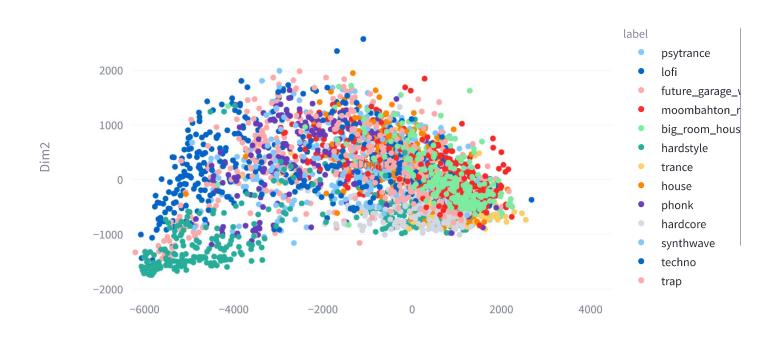
But why clustering?

Different from other music genres, EDM genres are highly overlap and blend (e.g. Tech House is a mix of Techno and House). By applying PCA, UMAP and t-SNE, we can see: 1. Which genres are naturally grouped together. 2. Which genres are distinct and separate. 3. Which genres have significant overlap, indicating subgenres.

localhost:8501 5/10 Select Dimensionality Reduction (You can try all three of them):

PCA

PCA Projection of EDM Genres



I selected PCA and it looks like a bag of Skittles just exploded on my screen.

This is actually a Principal Component Analysis projection, which is a fancy way of saying "Let's take our super high-dimensional EDM data and squish it down to just two dimensions, so our human brains (and eyes) can actually understand it." Each dot represents a song, and its position is determined by its compressed spectral features. Similar songs should, in theory, cluster together—but as you can see, real-world music isn't always so neatly categorized.

Those green blob in the bottem left are Ambient, which tends to have a distinct, tightly packed sound signature. Lofi (blue dots) also forms a cluster on the left, suggesting its tracks share a strong sonic identity.



Train a Genre Classification Model

So far, we've visualized our EDM genres using dimensionality reduction techniques like PCA. But here's the big question:

Can we teach a machine to recognize these genres automatically?

localhost:8501 6/10

I have a better question: why do we need to teach the machine?

Instead of just looking at those pretty clusters, models allow us to actually predict the genre of an unknown song based on its spectral characteristics. This model will learn from existing data and try its best to label new tracks accurately. And once we've got our model trained, we can push it even further—by identifying tracks that don't fit neatly into any category. These outliers could be experimental songs, hybrids, or misclassifications, offering insights into how genres evolve and intersect.

Click This To Train Random Forest Model



Accuracy: 0.93

precision recall f1-score support

ambient 0.94 0.92 0.93 37
big_room_house 0.95 1.00 0.98 42
dnb 0.92 0.92 0.92 39
dubstep 0.90 0.90 0.90 39
future_garage_wave_trap 0.89 0.93 0.91 42
hardcore 0.91 0.98 0.94 41
hardstyle 0.93 0.95 0.94 43
house 0.95 0.88 0.91 42
lofi 0.93 0.93 0.93 40
moombahton_reggaeton 0.82 0.95 0.88 39
phonk 0.93 0.93 0.93 43
psytrance 0.96 0.77 0.86 35
synthwave 0.85 0.87 0.86 39
techno 0.95 0.95 0.95 38
trance 1.00 1.00 1.00 40

accuracy 0.93 640 macro avg 0.93 0.92 0.92 640 weighted avg 0.93 0.93 0.92 640

trap 1.00 0.90 0.95 41

localhost:8501 7/10

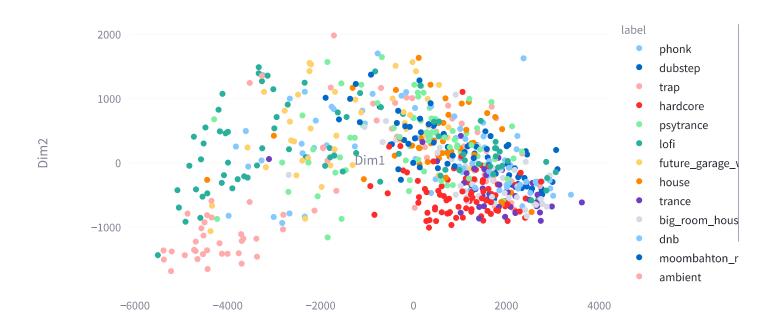
Total Detecting Imposters





0.10

Potential Genre Imposters



You should be seeing a chart that looks like the PCA result, but with less Skittles.

In this chart, we're detecting imposters—tracks that don't quite fit into their assigned genre. Think of it like a costume party where everyone is dressed as their favorite genre, but some guests clearly didn't read the dress code.

Each dot represents a track, and their positions are determined by their spectral characteristics. Most dots cluster together in genre-based groups, but some tracks are way off the expected region. These misfits, the ones that stray too far from their genre's usual neighborhood, are flagged as potential outliers (imposters).

Why are they imposters? Some producers love bending the rules. A lofi track with unusually high energy or a hardcore track with softer elements might show up here. Some tracks sit between genres, like a house track with techno influences or a dubstep track with trap beats. These confuse the model because they

localhost:8501 8/10

share features with multiple genres. Most importantly, the model isn't perfect. Sometimes, a song that belongs to one genre gets misclassified into another, making it an imposter in its assigned group.

The presence of imposters reminds us that genres aren't rigid categories. They evolve, blend, and borrow from each other. What seems like an imposter today might define a new subgenre tomorrow!

In All Words...What Did We Learn?

EDM has never been about fitting into neat little boxes—it thrives on evolution, fusion, and defiance of strict classification. Throughout this data-driven journey, we've uncovered the hidden patterns in sound, from spectral fingerprints that give each genre its unique identity to machine learning models that try (and sometimes hilariously fail) to categorize music into rigid labels. But in the end, what we've really revealed is the illusion of fixed genres. Genres in EDM are more like guidelines than rules—overlapping, borrowing, and constantly reshaping themselves. What sounds like House today might be labeled Future Garage tomorrow, and what was once dismissed as "just noise" could define the next big movement. Producers can use spectral insights to refine their sound, nudging their tracks toward a particular vibe—but that doesn't mean the music itself follows the rules. Listeners, too, can explore EDM beyond labels, discovering that their favorite "Techno" track might have more in common with Psytrance than they thought. But perhaps the most important takeaway is that music will always resist being neatly categorized. Just like the imposters in our visualizations, innovation thrives in the unexpected—those tracks that blur genres, challenge norms, and redefine what EDM can be.

References

Banitalebi-Dehkordi, Mehdi Amin. "Music Genre Classification Using Spectral Analysis and Sparse Representation of the Signals. https://arxiv.org/abs/1803.04652

Fedden, Leon. "Comparative Audio Analysis With Wavenet, MFCCs, UMAP, t-SNE and PCA." Medium, 20 Nov. 2017. https://medium.com/%40LeonFedden/comparative-audio-analysis-with-wavenet-mfccs-umap-t-sne-and-pca-cb8237bfce2f

Varma, Aastha. "Dimensionality Reduction: PCA, t-SNE, and UMAP." Medium, 14 July 2024. https://medium.com/%40aastha.code/dimensionality-reduction-pca-t-sne-and-umap-41d499da2df2

localhost:8501 9/10

localhost:8501 10/10