Project Proposal

**Hila Shoshan** ████████

Dataset Description

The data contains music from the following genres:

blues, classical, country, disco, hiphop, jazz, metal, pop, reggae and rock.

There are several data bases that refer to the same music, and I will be able to work with any of them.

- <u>Audio files dataset</u> - A collection of 10 genres with 100 audio (wav) files each, all having a length of 30 seconds.
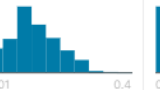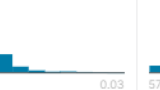 For example:

  < **metal.00002.wav** (1.26 MB)   ⬇ ⛶

  ▶ 0:00 / 0:30 ●──────────────── 🔊 ⋮
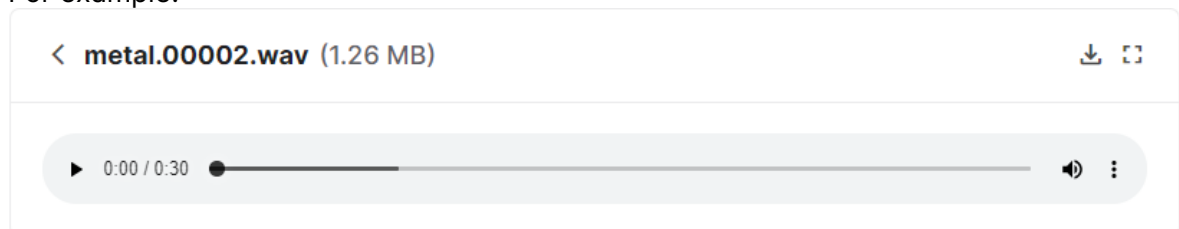
- <u>Images dataset</u> - A visual representation for each audio file. The audio files were converted to Mel Spectrograms.

  < **metal00002.png** (103.86 KB)
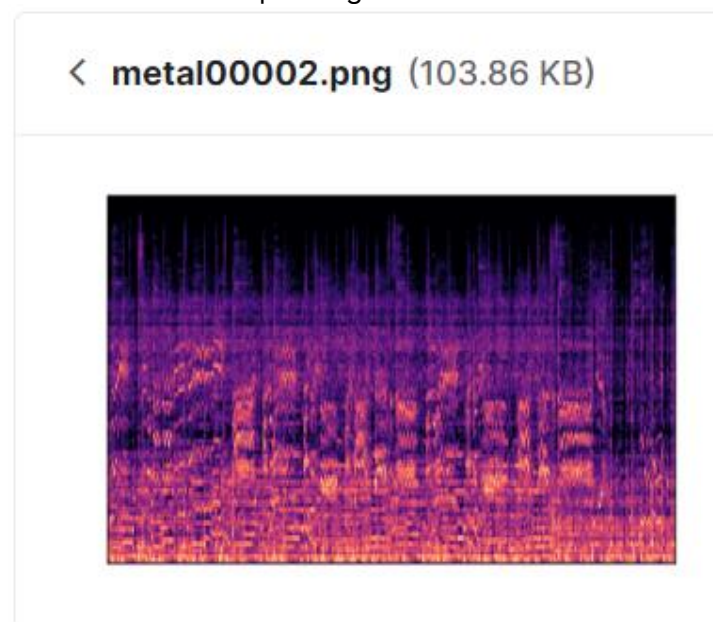
- <u>2 CSV files</u> - Containing features of the audio files. One file has for each song (30 seconds long) a mean and variance computed over multiple features that can be extracted from an audio file. The other file has the same structure, but the songs were split before into 3 seconds audio files (this way increasing 10 times the amount of data we fuel into our classification models).

⌶ filename	# length	# chroma_stft_mean	# chroma_stft_var	# rms_mean	# rms_var	# spectral_centroid...	⌶ label
1000 unique values	660k — 676k	0.17 — 0.66	0.04 — 0.11	0.01 — 0.4	0 — 0.03	570 — 4.44k	**10** unique values
blues.00000.wav	661794	0.35008811950683594	0.08875656872987747	0.1302279233932495	0.0028266964945942163	1784.165849538755	blues
blues.00001.wav	661794	0.3409135937690735	0.09498025476932526	0.09594780951738358	0.00237273913808167	1530.1766787460795	blues
blues.00002.wav	661794	0.36363717913627625	0.08527519553899765	0.17557041347026825	0.0027459163684397936	1552.8118647610036	blues
blues.00003.wav	661794	0.4047847092151642	0.09399903565645218	0.14109300076961517	0.006346346344798803	1070.1066149971282	blues
blues.00004.wav	661794	0.30852603912353516	0.0878409817814827	0.09152871370315552	0.0023033977748619318	1835.0042655120715	blues
blues.00005.wav	661794	0.3024562895298004	0.08753237873315811	0.10349363833665848	0.0039805080741643906	1831.993939606849	blues
blues.00006.wav	661794	0.29132798314094543	0.09398132562637329	0.14187414944171906	0.008803188800811768	1459.366471695765	blues
blues.00007.wav	661794	0.3079547882080078	0.09290287643671036	0.13182218372821808	0.005531143397092819	1451.6670658428209	blues
blues.00008.wav	661794	0.40887922048568726	0.0865124762058258	0.14241649210453033	0.0015071560628712177	1719.3689480431458	blues

The questions I would like to answer

The main question, for which I chose the dataset, is of course:
Is it possible to classify a song for its genre based on an audio segment from it?

The question can be divided into some more specific sub-questions like:

1. What is the technique that will bring the best results?

2. Will using the techniques we learned with the data from the CSV files only yield good results?

3. Will deep learning techniques, like CNN on the pictures data will predict the genre in a good way?

4. Are there types of songs that a model would have a very hard time distinguishing between?

5. Do genres that are not at all close to each other still have the same characteristics (for example in a certain part of the music)?

The techniques I will use

Adaboost, SVM, KNN, Decision trees, Clustering and CNN (deep learning).