

ML Project Proposal “Genre detection”

Project title: Genre detection with Spotify

Group members:

- Marcel Ciesla
- Rene Hörschinger
- Victoria Oberascher

Problem definition

Our goal is to build a model which can detect the genre of a song. So the input is a song, the output is a genre. With this model we can then detect all genres of a given playlist (from a user for example) and try to recommend songs that might suit the user's taste according to the genre.

Furthermore we like to investigate the features in detail, which ones are more important than others, which ones can be left out and are there other features we can derive from the given ones. We want to focus a lot on **data visualization, data partitioning, preprocessing** and **feature extraction** in order to not only build a model with a high detection probability, but also a model we fully understand with a workflow we can later use for different types of problems.

Dataset

There might be a dataset for this task, and we will of course use one in addition if it suits our problem statement well enough. But we want to focus on building our own dataset with the Spotify API. This allows us to get audio features for every song.

In our dataset, each sample is a song, and we want to annotate the genre by hand. Our idea is to use whole playlists from spotify (and self made playlists) with a very specific genre (Pop, Rock, Metal, Jazz, ...) so we can set the genres more or less automatically by saying “this playlist's genre is rock, annotate all tracks with rock”.

A sample response from the API for a specific track might look like this:

```

"segments": [
  {
    "start": 0.70154,
    "duration": 0.19891,
    "confidence": 0.435,
    "loudness_start": -23.053,
    "loudness_max": -14.25,
    "loudness_max_time": 0.07305,
    "loudness_end": 0,
    "pitches": [
      0.212,
      0.141,
      0.294
    ],
    "timbre": [
      42.115,
      64.373,
      -0.233
    ]
  }
],
"tatums": [
  {
    "start": 0.49567,
    "duration": 2.18749,
    "confidence": 0.925
  }
]
},
"track": {
  "num_samples": 4585515,
  "duration": 207.95985,
  "sample_md5": "string",
  "offset_seconds": 0,
  "window_seconds": 0,
  "analysis_sample_rate": 22050,
  "analysis_channels": 1,
  "end_of_fade_in": 0,
  "start_of_fade_out": 201.13705,
  "loudness": -5.883,
  "tempo": 118.211,
  "tempo_confidence": 0.73,
  "time_signature": 4,
  "time_signature_confidence": 0.994,
  "key": 9,
  "key_confidence": 0.408,
  "mode": 0,
  "mode_confidence": 0.485,
  "codestring": "string",
  "code_version": 3.15,
  "echoprintstring": "string",
  "echoprint_version": 4.15,
  "synchstring": "string",
  "synch_version": 1,
  "rhythmstring": "string",
  "rhythm_version": 1
},
"bars": [
  {
    "start": 0.49567,
    "duration": 2.18749,
    "confidence": 0.925
  }
],
"beats": [
  {
    "start": 0.49567,
    "duration": 2.18749,
    "confidence": 0.925
  }
],
"sections": [
  {
    "start": 0,
    "duration": 6.97092,
    "confidence": 1,
    "loudness": -14.938,
    "tempo": 113.178,
    "tempo_confidence": 0.647,
    "key": 9,
    "key_confidence": 0.297,
    "mode": -1,
    "mode_confidence": 0.471,
    "time_signature": 4,
    "time_signature_confidence": 1
  }
]
],

```