**Beskrivelse af slutprojekt**



|  |  |
| --- | --- |
| **Navn** | **Studienr** |
| Jacob Odgaard Hausted | 201510912 |
|  |  |
|  |  |

# Dataset

Det er valgt at gruppen i slutprojektet ønsker at arbejde med et datasæt hentet direkte via. Spotify’s API, ”*Spotify Tracks DB*”[[1]](#footnote-1). Datasættet indeholder sange som findes på spotify platformen, og disse findes beskrevet ved 16 features, hvortil nogle knytter sig sangens identifikation, eg. Navn, genre og artist. Resten beskriver forskellige aspekter af sangens *lyd* og *feel:*

Identifikation:

* Genre
* Artist name
* Track name
* Track ID
* Popularity

*Lyd* og *feel*:

* Acousticness
* Danceability
* Duration in ms
* Energy
* Instrumentalness
* Key
* Liveness
* Loudness
* Mode
* Speechiness
* Tempo
* Time signature
* Valence

Datasættet er fundet på kaggle og indeholder data for tusinde sange, hvortil det samtidig bør påpeges at dette kan udbygges da spotify’s API står frit tilgængeligt. De ovenstående features indgår alle i *Audio Features* objektet fra API’et, og det er netop 10000 af disse objekter fra 26 forskellige genrer der udgør datasættet.

Udover de tidligere benævnte features kan listen udvides, da API’et stiller et yderligere *Audio Analysis object* til rådighed. Heri vil der være mulighed for at opdele være sang i flere segmenter og få beskrevet hvert segments features(ovenstående for det givne segment) med tilhørende konfidens niveauer – altså et mål for hvor præcise featuresne er for det givne segment.

# Anvendelse:

Hovedidéen for datasættet vil være at prøve at prædiktere features, evt. genre og popularitet ud fra featuresne defineret i *lyd* og *feel.* Til netop dette ønsker gruppen også at udvide datasættet med *Audio Analysis* objektet, hvortil der vil blive mulighed for at teste sammenhængen imellem spotify’s confidens mål og fundne type I og II fejl.

1. https://www.kaggle.com/zaheenhamidani/ultimate-spotify-tracks-db [↑](#footnote-ref-1)