

Capstone Project Proposal
By
Harshavardhan Nilakantan

There are many different genres of music. When songs are classified to be of a certain genre, they are classified as such due to some similarity in their musical qualities - often pertaining to a structural or stylistic property contained in the song. While these cannot be easily quantified, they can be distinguished by a trained ear. Similarly, I have often felt that a lot of songs could potentially share a certain degree of similarity based upon certain quantifiable characteristics or descriptive information that they might share.

Given this idea, I would like to explore a data set containing audio properties such as popularity, time signature, instrumentality, tempo, danceability etc. Without any pre-definitions, I would like to see if they fall into any grouping based solely on these characteristics.

The dataset would be obtained from Kaggle. I would use various kinds of clustering techniques to explore this dataset, and expose any underlying structures.

I will then share the code and further documentation based on my analyses and understanding. More specifically, I will submit all Jupyter notebooks that I develop; I will work on a final written report and a presentation slide deck.