**Unsupervised Learning and Finding Similar Artists using Echonest Song Data**

David Reinke

Ricky Reyes

Dimitri Dimov

Program Contents:

EchoNestScraper.py

Unsupervised\_similar\_artists.py

myartists.txt

songdata.txt

This program uses unsupervised learning in order to predict similar artists using the **k-Means algorithm** in the Python sklearn library. Using a list of artists we created a collection of song data acquired from Echo Nest using the pyechonest library compatible with Python 3.5 at <https://github.com/vrangasayee/pyechonest/tree/python3>.

**EchoNextScraper.py** – the python script which reads in a list of artists and gathers song information for each artist from EchoNest and saves it to songdata.txt

**Unsupervised\_similar\_artists.py** – reads in the song data from songdata.txt and configures an unsupervised model on chosen features using the sklearn k-means algorithm in order to cluster the song data. Using a user interactive GUI, the user is asked to enter an artist and the script will query EchoNest for a song from the entered artist. We then predict similar artists by extracting the features of that song and fitting it to our model which assigns it to a cluster. We then return the 5 most frequent artists belonging to that cluster in our model.

**Configuration** :

Selected features: Energy, Loudness

KMeans clustering, clusters = 100,

**Potential Improvements:** the returned song data from echonest was not controlled. Perhaps gathering data for the most popular songs for each artist would provide a better representation of that artist. At user entry, has the same scenario. The one song we use to predict on which EchoNest provides us might not be a sufficient representation of an artist’s overall sound. Although we found that training our model using the energy and loudness features produced decent results, alternate algorithm configuration and features used to train our model might prove to produce more accurate results.