**Radiology**

**Team members:**

Pei Lian Liu – [esponavy@yahoo.com](mailto:esponavy@yahoo.com)

Long “Kevin” Trinh – [kevintrinh255@gmail.com](mailto:kevintrinh255@gmail.com)

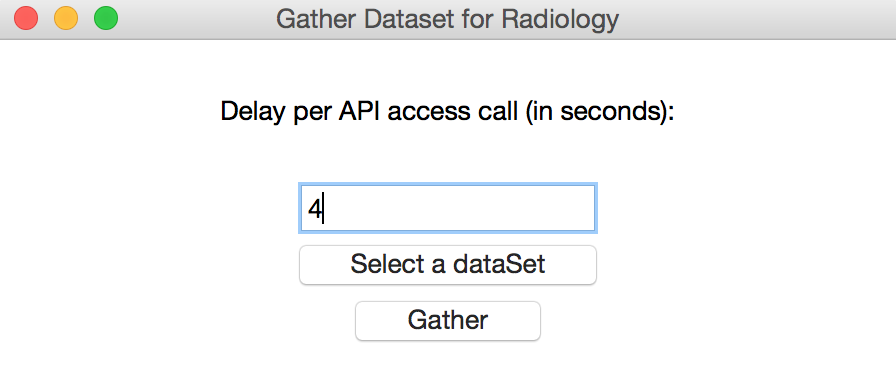
Peter Chen – peter.chen20@gmail.com

**Programming Language:** Python 2.7

**Our goal:** To find similar artists for a given artist using the unsupervised learning K-Means algorithm.

1. Gather data from Echonest

Due to the limitations given by Echonest’s API access rate, we decided to set up a loop to fetch a random song from their database at a specified interval. Recommended is more than 3 seconds because Echonest only allows you to call their API at 20 calls per minute.



We have provided datasets to test and use the main application located in the datasets folder.

*small\_dataset.txt contains 2,720 songs with the genres: classical, pop, rock, punk, and rap*

*huge\_dataset.txt contains 8,256 songs with the genres: classical, pop, rock, punk, and rap*

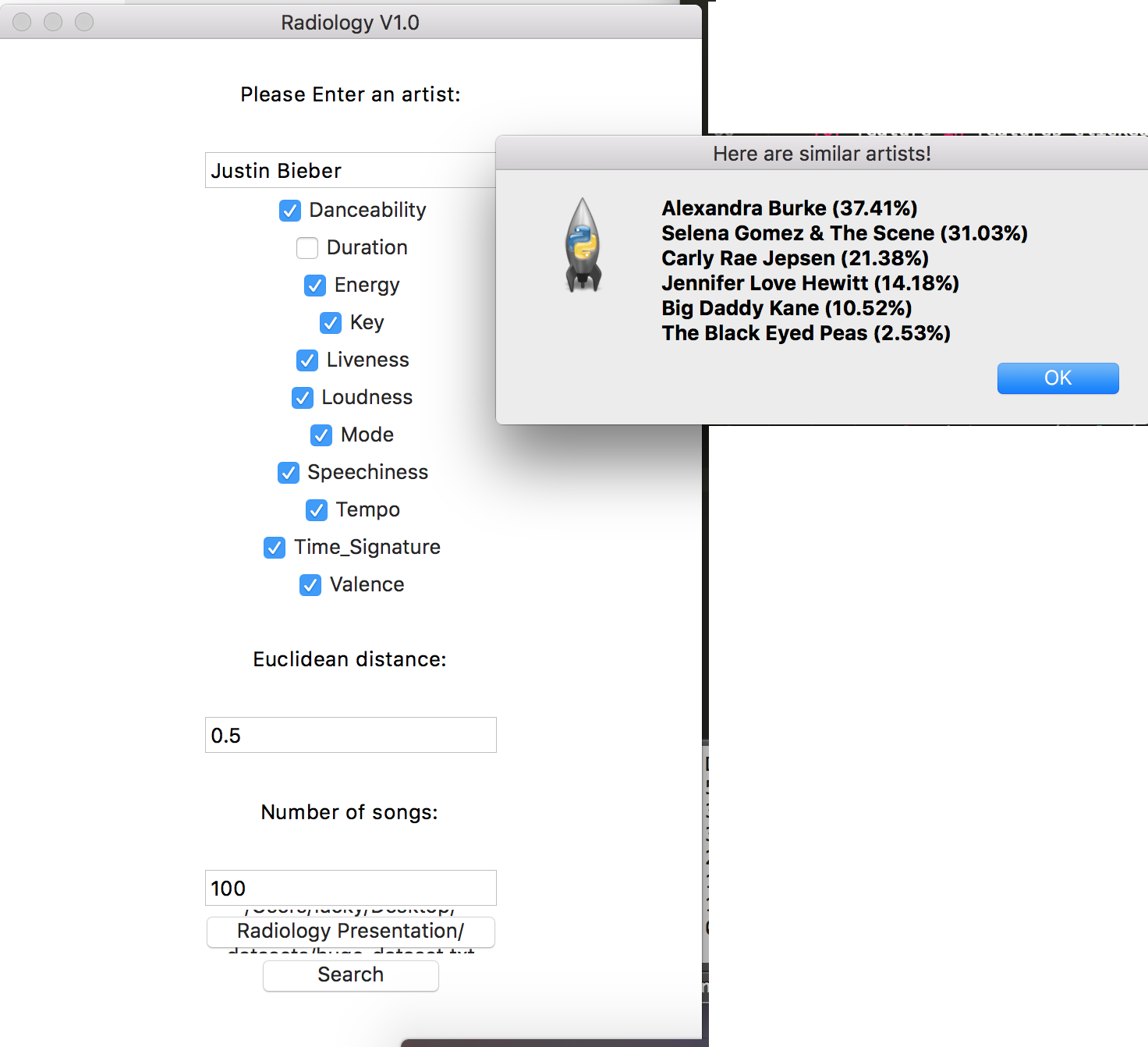
*classical\_only.txt contains 1,430 songs with the genres: classical*

*empty\_dataset.txt contains 0 songs with the genres: ---*

*jazz\_chinese\_opera\_light\_music\_punk\_latin.txt contains 2,418 songs with the genres: jazz, chinese opera, light music, punk, and latin*

*chinese\_opera\_meditation\_punk\_pop\_rap\_kpop.txt contains 1,273 songs with the genres: chinese opera, meditation, punk, pop, rap, and k-pop*

1. Use application to find similar artists in dataset.



The results that are shown after hitting “Search” are similar artists only found in the dataset.

We obtain the features from the number of songs of the specified artist. After that, we fit our K-Means model (number of clusters is 5) with features from songs in the datasets.

For each song of the specified artist, we predict the clusters at which they fall under. For every cluster they fall under, we compute the Euclidean distance (of the feature sets) of that song to every song in the cluster. Those that fall under a certain specified threshold will be added to the list of similar artists returned.

For convenience, we’ve limited the results to 5 (they are sorted to display the most highly matched) and display a similarity rating which is the actual Euclidean distance divided by the input of Euclidean distance.

In conclusion, the overall of our project works well and our program would be better if we have more time to working on it. We like working in a team because everyone has their own strengths so that we can share the experience and figure out the problems easily and efficiently.