

JukeBox

1

Content Aggregation

- Go through YouTube and get links for top 25 songs
- Ensure each song obtained has a viable chorus

2

Crowd-Sourcing I:

- Upload all these links onto Mechanical Turk, asking users to hum the song back on Vocaroo
- After HITs are complete, we will compile all of these audio files on Vocaroo on a .csv file

3

Quality Control:

- We will use this .csv file for our second HIT, in which we use the crowd to perform quality control
- This HIT will include the data we have collected, as well as positive and negative "gold standard" control tests to ensure the quality of verifiers

4

Feature Extraction

- For each of the verified samples, we webscrape the links to download all the quality controlled links
- Using this .mp3 files, we will run our feature extraction algorithm

5

ML Training:

- Using these feature extracted audio samples as a data set, we will train our ML algorithm
- It will associate audio features with our song title label

6

User Product:

- We will create a user platform that allows students to record their own voices, submit
- After submission, we will standardize, normalize the sample
- Our humming recognition algorithm will run, returning the user with a predicted song title