

Project Name: Accelerando

Team:

Name	Email
Chongyang Chen	cc5502@nyu.edu
Dongjie Fan	df1676@nyu.edu
Sichen Tang	st1671@nyu.edu

Purpose:

Accerlerando aims to build a Music & Artist Recommendation System.

Step:

For our final code, since we want our code to be reproducible, we generated two python script that can be conducted efficiently in dumbo. The 1st python script **dumbo1.py** aims at recommending music for users, as we mentioned before, this script includes three functions, **evaluation** gives us evaluation score of hyper-parameter **rank** for our ALS model, **bestRanking** selects the best **rank** for our ALS model, **matchsong** helps us to print out the exact song name for the predicted songid, **matchArtist** print out the Artist name from the artistid.

In our python script main method we read our datasets using *SQL* query, call the functions and then use **ALS.trainImplicit** function in **pyspark.mllib.recommendation** package to predict Top 3 recommendations for one given user.

The **dumbo2.py** aims at recommending artist for users, we also include the three functions, the only difference from **dumbo1.py** is that we generated a new dataset for artist with 'user - artist - counts' format, then run the same steps as for music recommendation.

In this code package, we also include output for these two Python scripts. Also we upload the datasets. Since it takes a while to run the whole dataset, it might be better to use the test dataset(*id_train_triplets.csv*) to check how the scripts work.