Authors: J. Hamski
Document: Final Project Proposal
Date: July 10, 2016

DATA 643 (Recommender Systems) Final Project Proposal

Business Goal:

The business goal of my final project is to build a recommendation system using Apache Spark on the Google Cloud platform. I aim to provide musician recommendations using the Last.fm dataset from 2006. This recommender system will have a basic web front end for interaction with the results.

Tech Stack:

My final project will run Apache Spark on Google Compute Engine. In addition, it will utilize Google App Engine and Google Cloud SQL. I am not a systems engineer or backend developer, but luckily there exists a steb-by-step tutorial for building a recommender system using Google Cloud:

Using Machine Learning on Compute Engine to Make Product Recommendations

Instructions are contained in a GitHub repository:

Spark Recommendation Engine

Note that my project will be using the platform architecture used in the tutorial, but will be departing from the tutorial on dataset used and the data analysis performed.

Tutorial Steps:

1. Setup a Spark cluster.
2. Setup a simple Google App Engine website.
3. Create a Cloud SQL database with an accommodation table, a rating table and a recommendation table.
4. Run a Python script on the Spark cluster to find the best model.
5. Run a Python script making a prediction using the best model.

6. Saving the predictions into Cloud SQL so the user can see them when displaying the welcome page.

Data Source:

I will continue to use the Last.fm dataset that contains:

- Listeners and a count of listens for particular artists
- LIsteners social graph
- Listener-generated tags for artists

Audience:

The audience for this recommender system will be people who want to music recommendations as if it were still 2006 - since that's when the dataset records were created!