What should I watch next?

A Movie Recommender Engine by Kevin Perlas

Chances are, you have already seen a recommender engine. If you been on YouTube, there is a “Recommended for You” section on the homepage. Netflix has recommendations when you open its app and after the end of viewing every movie. Amazon will bombard you with items “because you viewed X,” “because you bought Y”, or “Inspired by your trends.” Other industries use recommender engines outside of selling you things. Waze uses them for intelligent navigation systems; IBM uses them for traffic control systems, and GE uses them to find the best routes to save energy.

# **What is a Recommender Engine?**

Think of a highly sophisticated salesperson that knows you, your taste, and your style. They can also make intelligent decisions about what recommendations would benefit you the most. A very thorough definition by F.O. Isinkaye is found below:

Recommendation Engines are information filtering systems that deal with the problem of information overload by filtering vital information fragment out of a large amount of dynamically generated information according to user’s preferences, interest, or observed behavior about the item. Recommendation Engines / Recommender system has the ability to predict whether a particular user would prefer an item or not based on the user’s profile.[[1]](#footnote-1)

# **Content-Based Filtering**

When a friend asks you for a movie recommendation, you would probably ask what kinds of movies they like. From there, you could think of a few titles that are similar to the things they’ve liked in the past. This process, of recommending content based on its attributes, is basically **content-based filtering.** It may not be as trendy as its machine learning counterpart **collaborative-based filtering**, but it has many advantages and is used by popular applications such as Pandora’s Music Genome Project.

The Music Genome Project doesn’t compare your preferences or shopping habits with those of others. It doesn’t care whether the artists you like (or suggests) are already popular or just starting out. It levels the playing field by relying exclusively on the unique musical quality of each song. By painstakingly analyzing each song across 400 musical traits, we’ve make it easy for music lovers to discover new songs and artists.

-Tim Westergren, Co-Founder Pandora (former CEO)

# **Collaborative-Based Filtering**

There is a lot sophisticated math and machine learning techniques involved, but the concept behind it is straightforward enough: It’s based on the idea that people who share an interest in certain things will probably have similar tastes in other things as well. When you go shopping on Amazon and see “Customers Who Bought This Item Also Bought,” or “Users like you also liked,” that is **collaborative-based filtering.**

# **MovieLens 20M Dataset**

This is the dataset used in this project. It is a stable benchmark dataset. 20 million ratings and 465,000 tag applications applied to 10,000 movies by 138,000 users. Includes tag genome data with 12 million relevance scores across 1,100 tags.

# **Python Packages Used**

* Scikit-Learn
  + Machine Learning in Python
  + <http://scikit-learn.org/>
* Surprise
  + Python Scikit for recommendor systems
  + <http://surpriselib.com/>

# **Want to see more?**

Check this project on Github: <https://github.com/kevinperlas/my_workspace>

1. Recommendation systems: Principles, methods and evaluation by [F.O.Isinkaye](http://www.sciencedirect.com/science/article/pii/S1110866515000341#!), et al <<http://www.sciencedirect.com/science/article/pii/S1110866515000341>> [↑](#footnote-ref-1)