Primary Components of a recommender System are

1. Candidate generation
   1. In this first state, the system starts from huge dataset and generates a smaller subset of data. Example in Youtube reduces large amount of videos down to hundreds or thousands. This is first stage of Recommendation. two common candidate generation approaches:
      1. Content-based Filtering
         1. Uses similarity between items to recommend items similar to what user likes
         2. Ex: if User watches dog videos, then the system can recommend cute animals videos to that user.
      2. Collaborative Filtering
         1. Recommends items based on the behavior or preferences of similar users.
         2. Ex: If used A is similar to user B, and user B likes video 1, then the system can recommend video 1 to user A.
      3. Hybrid Recommender
         1. Hybrid recommender systems combine collaborative filtering (analyzing user behavior and preferences) and content-based filtering (focusing on item characteristics) to provide more accurate and diverse recommendations.
2. Scoring
   1. This model scores and ranks the candidates in order to select the set of items (on the order of 10) to display to the user. This is subset of Candidate generation
3. Re-ranking
   1. This is final ranking which takes into account dislikes or likes of newer content. Re-ranking helps to ensure diversity, freshness and fairness.

To determine the degree of similarity, most recommendation systems rely on one or more of the following:

1. Manhattan

Also, know as city block distance, or taxicab geometry wherein the distance is measured between two data points in Grid-like path. As shown below. Mainly used where high dimensionality in the data.

A grid with a red line

Description automatically generated

1. Euclidean

This is the shortest distance between 2 data points in the plane. Mainly used for the smaller dimensionality problems.

A diagram of a person walking on a line

Description automatically generated

1. Cosine Distance

Also known as Cosine Similarity is used to find similarities between two data points. Cosine similarity is given by Cos θ, and cosine distance is 1- Cos θ. Mainly used in the Collaborative Filtering based recommender systems to offer future recommendations

A graph with a line and a point

Description automatically generated