**Movie Recommendation System**

Types of Recommendation:

* Content Based: Rom-Com watchers should be recommended Rom-Coms, we can base this on tags.
* Collabarative-filtering-based: If we find user A and B to be similar. Recommendation works like if A watches Inception, we recommended B to watch inception.
* Hybrid

This project is a content based recommendation system.

Flow:

Data → Preprocessing → Model building → make a website(?) → Deploy the website(?)

Make the data to be:

Movie id

Title

Tags → from all relevant columns

**Vectorization:**

1. We can go based off of percentage of common keywords in the tags.
2. Instead, let’s convert the tags into vectors

We can to do text → vector.

Text vectorization

I’ll use bag of words(we also have df idf, word to vec)

Bag of words:

Concat all the tags

Pull out the 5000 most common words from the large text

w1 | w2 | w3 | w3 | …. | w5000

Movie1 5 3 0 2 1

Movie2 2 2 6 1 0

.

.

.

Movie5000 . . . . .

Shape = 5000x5000

Each row is now considered a vector

Fetch the closest 5 vectors to the user

Lesser words the better → reduce the dimensionality

Do not consider the stop words → or and to from it is are that a

We calculate the Cosine distance between vertices(angle)

Distance is inversely proportional to similarity