**Type of recommendation system**

Content based: recommend on the type of content one view (tags) and tags similarity. (content similarity)

Collaborative based (rating based): recommend on the interest of the user (user similarity)

Hybrid based: combination of both.

Step 2

We did the data processing in which we formatted the data took the important column converted it in the form of tag (processing is done to convert list to string, drop na rows and concatenation etc).

Then the vectorization of data took place which was done by Count Vectorizer class of sklearn which convert text to vector which then can be used for recommendation and closely related movies.

But we should also remove the words like love, loving, loved etc. so we should do steaming but as we researched more we came to see the drawbacks of steaming so now we switched to wordnet lemmatize

What we

When we have large dimension data we use cosine distance

Now we calculate distance of every movie ie a vector point with every other movie

Cosine similarity **measures the similarity between two vectors of an inner product space**. It is measured by the cosine of the angle between two vectors and determines whether two vectors are pointing in roughly the same direction. It is often used to measure document similarity in text analysis.

Now we will find the closest 5 movie to the entered movie for that we have to find the index of the entered movie in our cosine similarity vector

We use the enumerate function because it creates a map with the key as the index because we cant mess that up cause that’s keeping track of the movies (ie similarity matrix [0] is a array in which 0 shows the distance of 0th movie with itself same goes on)

We have to also mention that the sorting in this map should be done on the basis of cosine similarity and not keys.