**Recommendation System**

**Content based**  
Content based recommendation system recommends content based on similarity of content. In this tags are created, and based on similarity of tags, recommendation is done.

[ genre, keywords, overview, title, cast, crew (director) ]

[popularity, year]

**Collaborative filtering based**Collaborative filtering based recommendation system recommends content based on user’s interest and similarity.

Let’s say there are two users A and B, and by their behavior we know that they are very similar (their similarity score is 0.9). We know this as they watch same/similar movies and their rating on same movies is very similar.   
Now if the user A watches a new movie and gives good rating, that movie will be recommended to user B as we know the user’s have a high similarity rate.

It is used in Facebook’s react to a post system.

**Hybrid**Hybrid recommendation system is a mixture of both content based recommendation system and collaborative filtering based recommendation system.

**Project Flow**

Dataset 🡪 Pre processing on data (converting the data as per our needs) 🡪 make ML model for recommendation 🡪 convert the model into a website 🡪 deploy the website

**Dataset**

TMDB: <https://www.kaggle.com/datasets/tmdb/tmdb-movie-metadata>

**Creating ML model**

1. Import dataset
2. Convert it into a dataframe
3. Pre-process the dataframe [select important columns, some processing on the columns (convert columns in proper format, pick top 3 casts from cast column, pick director from crew column), merge all the columns into one tags column]
4. Convert tags column to a vector by text-vectorization using the technique of bag of words.

[concat all the tags, get top 5000 words based on frequency, now get frequency of each word in all the movie tags, this frequency array is a vector (don’t include stop words)]

1. Find cosine distance (angle) between two vectors.