# MOVIE-RECOMMENDATION SYSTEM

PROJECT REPORT
BDM [3014]

GROUP - E
1) Kamaldeep Kaur Dhami
2) Kirandeep Kaur
3) Anirudh Siddula
4) Bibek Shah Shankhar
5) Davinderjit Singh

DECEMBER 9, 2021

# TOPIC – MOVIE RECOMMENDATION SYSTEM

## What is a recommender system?

A recommender system is a simple algorithm whose aim is to provide the most relevant information to a user by discovering patterns in a dataset. The algorithm rates the items and shows the user the items that they would rate highly. An example of recommendation in action is when you visit amazon and you notice that some items are being recommended to you or when Netflix recommends certain movies to you. This engine makes suggestions by learning and understanding the patterns in your watch history (let's say) and then applies those patterns and findings to make new suggestions.

There are 3 types of recommendation systems.

1. **Demographic Filtering:** The recommendations are the same for every user. They are generalized, not personalized. These types of systems are behind sections like "Top Trending".
2. **Content-based Filtering:** These suggest recommendations based on the item metadata (movie, product, song, etc). Here, the main idea is if a user likes an item, then the user will also like items similar to it.
3. **Collaboration-based Filtering:** These systems make recommendations by grouping the users with similar interests. For this system, metadata of the item is not required.

In this project, we are building a Content-based recommendation engine for movies.

## About Movie Recommendation Project

In this machine learning project, we build a recommendation system from the ground up to suggest movies to the user based on his/her preferences.

Project Dataset

There are several datasets available to build a movie recommendation system. But for this project, we are going to use a dataset that contains the metadata (cast, crew, budget, etc..) of the movie.

## How to build a Movie Recommendation System using Machine Learning

The approach to build the movie recommendation engine consists of the following steps.

1. Perform Exploratory Data Analysis (EDA) on the data
2. Build the recommendation system
3. Get recommendations

- Perform Exploratory Data Analysis (EDA) on the data
- The dataset contains one CSV file movies, it contains the information like name and id of the movie, budget, languages in the movie that has been released, etc.
- First, we loaded the movie dataset using pandas.
- We only need the keywords, genres, cast, and director columns of the dataframe. So, we created a new list of important columns 'features'.
- Then we check if there are any null values present in the dataset, it resulted TRUE so we clean and preprocess the data by filling any missing value with an empty string.
- We made a function to combine the values of the important columns into single string.
- We Created a new column called combined_features with keywords, cast, genres & director
- Our movie recommendation engine works by suggesting movies to the user based on the metadata information. The similarity between the movies is calculated and then used to make recommendations. For that, our text data should be preprocessed and converted into a vectorizer using the CountVectorizer. As the name suggests, CountVectorizer counts the frequency of each word and outputs a 2D vector containing frequencies.

- There exist several similarity score functions such as cosine similarity, Pearson correlation coefficient, etc. Here, we use the cosine similarity score as this is just the dot product of the vector output by the CountVectorizer.

- We made helper functions to get title from the index, index from title.

- User will enter the movie.

- Get the index of the movie using the title.

- Get the list of similarity scores of the movies concerning all the movies.

- Enumerate them (create tuples) with the first element being the index and the second element is the cosine similarity score.

- Sort the list of tuples in descending order based on the similarity score.

- Get the list of the indices of the top 5 movies from the above sorted list. Exclude the first element because it is the title itself.

- Map those indices to their respective titles and return the movies list.

- At last, we print the list of movies along with director's name.

## Summary

In this machine learning project, we build movie recommendation system. We built a content-based recommendation engine that makes recommendations given the title of the movie as input.