Presentation on:-

CONTENT-BASED MOVIE RECOMMENDATION SYSTEM

By

DEEP SIKDAR

ANURAG GHOSH

ANKOOR DAS

SIDDHARTHA CHAKRABORTY

CONTENTS : –

PROJECT OBJECTIVE

Objective:-

- **Given :** A movie details dataset taken from Kaggle (contains training and test data).
- **Goal :** To Recommended a list of 20-30 movie names according to the search movie name .

PROJECT SCOPE

Recommender system are information filtering tools that aspire to suggest a list of movies.

Movie Recommendation systems provide a mechanism to assist users in selecting with similar interests.

- This makes Recommender system essentially a central part of OTT websites and e-commerce applications.
- Currently the industry is trying to integrate various advanced recommender systems which work on group recommendation or POI.

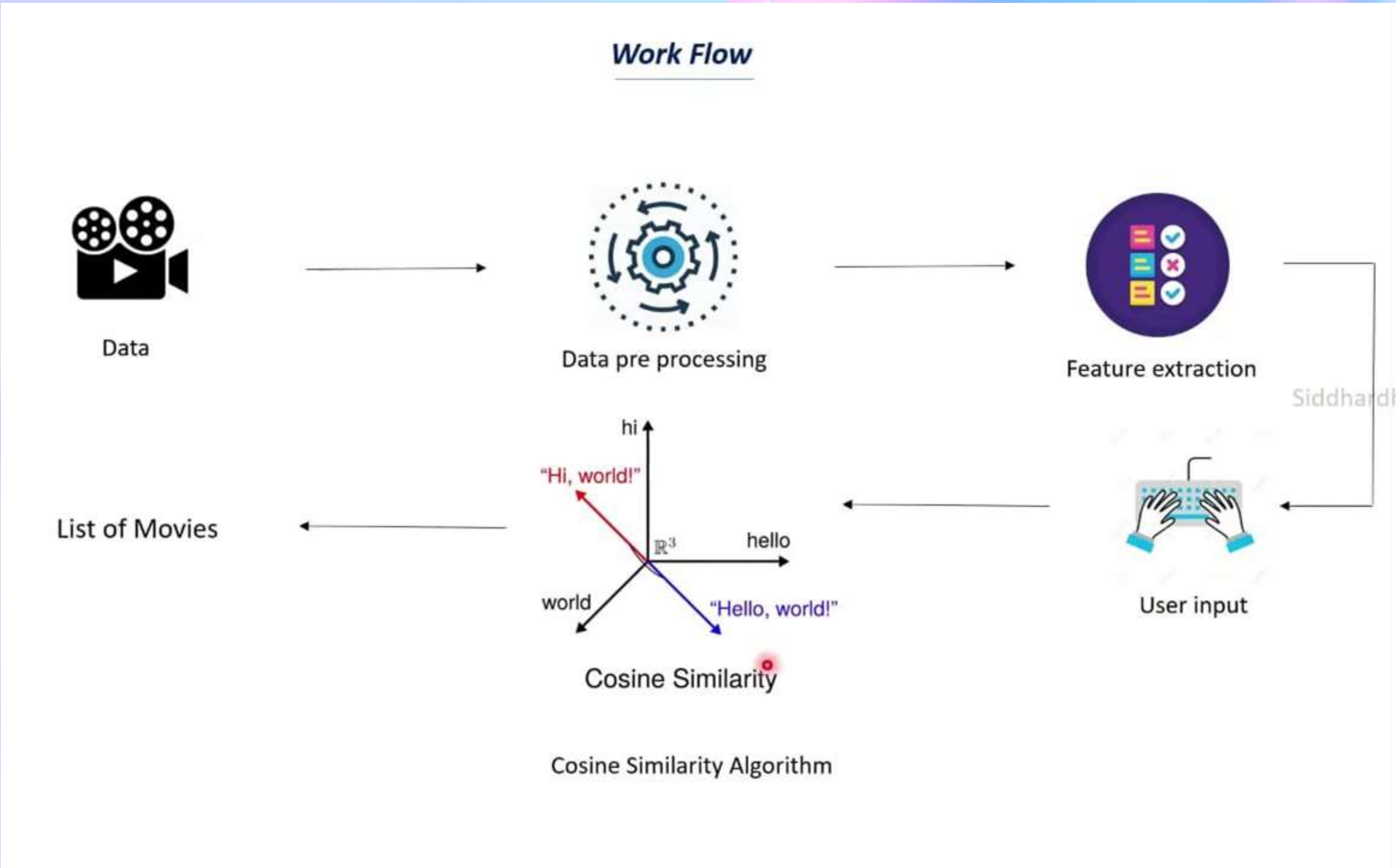
WHAT IS A RECOMMENDATION SYSTEM?

- **A movie recommendation system** is a fancy way to describe a process that tries to recommended your preferred items based on your or people similar to you.
- **Why exactly do we need Recommendation Systems?**
- From a user's perspective, they are catered to fulfill the user's needs in the shortest time possible. For example, the type of content you watch on Netflix or Hulu. A person who likes to watch only *Korean drama* will see titles related to that only but a person who likes to watch *Action-based* titles will see that on their home screen.
- From an organization's perspective, they want to keep the user as long as possible on the platform so that it will generate the most possible profit for them. With better recommendations, it creates positive feedback from the user as well.

METHODOLOGY

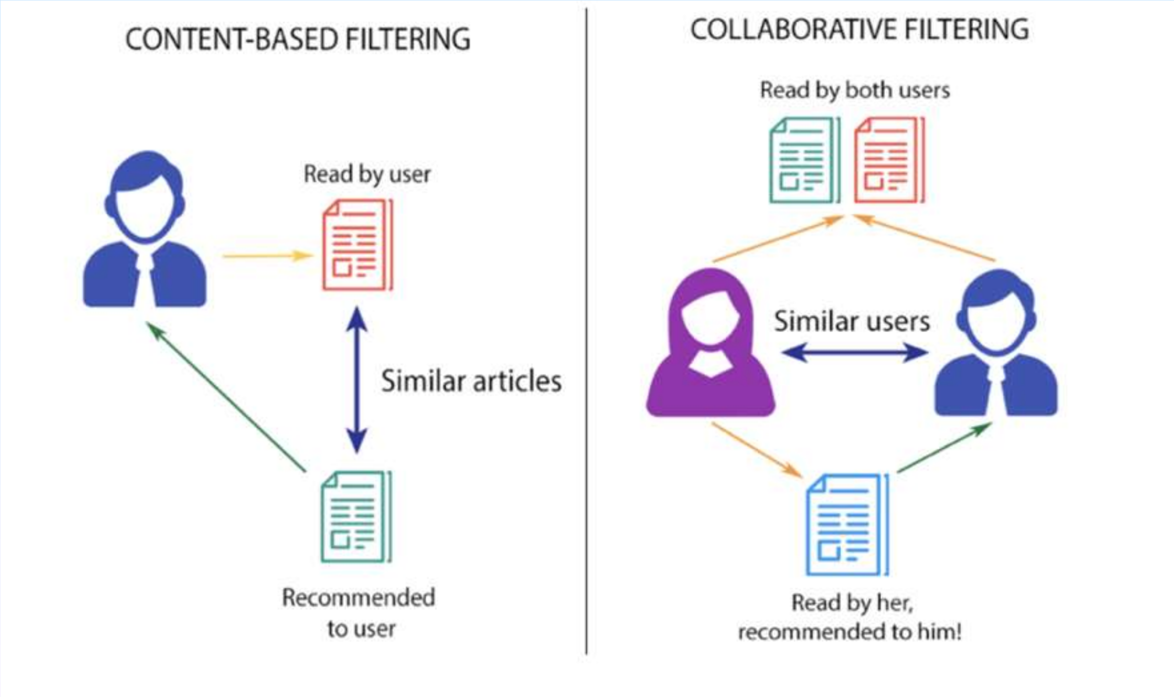
- ❑ **Collect and preprocess data**
- ❑ **Data Analysis**
- ❑ **Feature extraction**
- ❑ **Similarity calculation**
- ❑ **Movie recommendation**
- ❑ **Improvements in the project**

WORK-FLOW


Work Flow
Data
Data pre processing
Feature extraction
User input
hi
"Hi, world!"
hello
world
"Hello, world!"
Cosine Similarity
List of Movies
Cosine Similarity Algorithm

TYPE

- ➔Content-based filtering
- ➔Popularity based filtering
- ➔Collaborative filtering



DATA DESCRIPTION

A movie recommendation system requires a dataset of movies and their associated metadata, such as movie titles, genres, release dates, ratings, and user reviews. This metadata can be used to build a model that can suggest movies to users based on their preferences and past viewing history.

Here are some important data attributes that are commonly used in movie recommendation systems:

1. Movie title - The name of the movie.
2. Genre - The category or type of movie, such as action, comedy, drama, etc.
3. Release date - The date when the movie was released in theaters or on video.
4. Director - The name of the person who directed the movie.
5. Cast - The names of the actors who starred in the movie.
6. Rating - The rating of the movie, such as G, PG, PG-13, R, or NC-17.
7. ratings - The rating that users have given to the movie, usually on a scale of 1-5 or 1-10.
8. User reviews - Comments and feedback provided by users who have watched the movie.

Other attributes may include plot summary, duration, production company, language, country of origin, and more. This metadata can be used to build a model that can analyze a user's viewing history and suggest movies that are similar in genre, theme, or other attributes.

It is important to have a large and diverse dataset to ensure accurate recommendations. The data should also be regularly updated to keep up with new releases and changing user preferences.

DATA PREPROCESSING

Data preprocessing is an important step in building a movie recommendation system. Here are some common preprocessing steps that can be applied to movie data:

1. Data cleaning - This involves removing any irrelevant or duplicate data, and filling in any missing data values. For example, if a movie is missing a rating, the missing value can be replaced with the average rating of all movies in the dataset.
2. Feature engineering - This involves creating new features from the existing data to improve the accuracy of the recommendation system. For example, new features can be created based on the director, cast, or genre of the movie.
3. Normalization and scaling - This involves transforming the data to make sure it is on a similar scale. This can be important for some machine learning models, such as clustering or nearest-neighbor methods.
4. Dimensionality reduction - This involves reducing the number of features in the dataset. This can be useful when dealing with large datasets that contain many features, or when the features are highly correlated. Techniques such as principal component analysis (PCA) can be used to reduce the number of features.
5. Data sampling - This involves selecting a subset of the data for analysis. This can be useful when dealing with large datasets that may be difficult to analyze in their entirety.

Once the data has been preprocessed, it can be used to train a machine learning model that can provide recommendations to users based on their preferences and viewing history.

MODELS USED

TF-IDF

TF-IDF stands for Term Frequency-Inverse Document Frequency. It is a commonly used technique for information retrieval and text mining to represent the importance of each word in a document or corpus.

TF-IDF takes into account the frequency of a term in a document and the frequency of the term in the corpus as a whole. The term frequency (TF) of a term in a document is the number of times the term appears in the document. The inverse document frequency (IDF) of a term is a measure of how common or rare the term is in the corpus. The IDF is calculated as the logarithm of the total number of documents in the corpus divided by the number of documents in the corpus that contain the term.

The TF-IDF score of a term in a document is calculated by multiplying the term frequency of the term in the document by the inverse document frequency of the term. The result is a numerical score that reflects the importance of the term in the document and the corpus as a whole.

TF-IDF is used in many natural language processing tasks, including text classification, information retrieval, and recommendation systems. It is a powerful tool for identifying important terms in a document and for comparing the similarity of different documents based on their content.

MODELS USED

Cosine Similarity Algorithm

Cosine similarity is a measure of similarity between two non-zero vectors of an inner product space. It is a widely used similarity measure in recommendation systems, natural language processing, and information retrieval.

Cosine similarity measures the cosine of the angle between two vectors, where the vectors are represented as n-dimensional vectors in a multi-dimensional space. If the angle between two vectors is small (close to 0 degrees), then the cosine of the angle is large, indicating that the vectors are similar. If the angle between two vectors is large (close to 90 degrees), then the cosine of the angle is small, indicating that the vectors are dissimilar.

In recommendation systems, cosine similarity can be used to calculate the similarity between two items or between an item and a user's preferences. The cosine similarity between two items is calculated as the dot product of the two items' feature vectors, divided by the product of the magnitudes of the two vectors.

Cosine similarity is a popular similarity measure because it is easy to calculate, works well with sparse data, and is not affected by the magnitude of the vectors. It is often used in combination with other similarity measures in recommendation systems to provide more accurate and personalized recommendations.

FUTURE SCOPE OF IMPROVEMENTS

- Incorporating more complex features
- Hybrid recommendation systems
- Personalization
- Dynamic updates

CONCLUSION

- ★In conclusion, a content-based movie recommendation system is a valuable tool for recommending movies to users based on their past movie preferences.
- ★The system can be further improved by using techniques such as natural language processing, similarity calculation, and user feedback. The content-based movie recommendation system can enhance user experience, increase user engagement, and improve the revenue of movie recommendation platforms.
- ★ Overall, the content-based movie recommendation system is an effective and efficient way to recommend movies to users, and its implementation can benefit both users and the movie industry.

THANK YOU

Asansol Engineering College

- Deep Sikdar(deepsikdar23@gmail.com)
- Ankoor Das(ankoor2002@gmail.com)
- Siddhartha Chakraborty(siddhartha.ukh@gmail.com)
- Anurag Ghosh(anurag0862@gmail.com)