Content-Based Movie Recommender System

(Group 30)

> **Prerak Panwar: Roll number 40, Section 2**
> **Yi Hsiang Hung: Roll number 13, Section 1**
> **Jai Saxena: Roll number 43, Section 1**

1. Introduction:

In today's online movie platforms, personalized suggestions are important for making users happy and interested. This project is about making a smart system that suggests movies using machine learning. It looks at what users like, details about movies, and past data to predict and suggest movies each user might enjoy. This will make watching movies more fun and enjoyable for everyone.

This project aims to develop a content-based movie recommender system that leverages content similarity to recommend movies to users based on their input. We will utilize a dataset containing 5000 movie details obtained from Kaggle. The recommendation system will be built using natural language processing techniques, specifically focusing on vectorization, cosine distance calculation, stemming, and building a recommender function.

2. Dataset Overview:

The dataset comprises information about 5000 movies, including attributes such as title, genre, keywords, director, cast, etc. This dataset will serve as the foundation for our content-based recommendation system.

Link: https://www.kaggle.com/datasets/tmdb/tmdb-movie-metadata

3. Types of Recommendation Systems:

We will focus on developing a content-based recommendation system for this project. However, it's worth mentioning that there are other types of

recommendation systems such as collaborative filtering and hybrid systems that combine multiple approaches. For this project, we will specifically emphasize content-based recommendations.
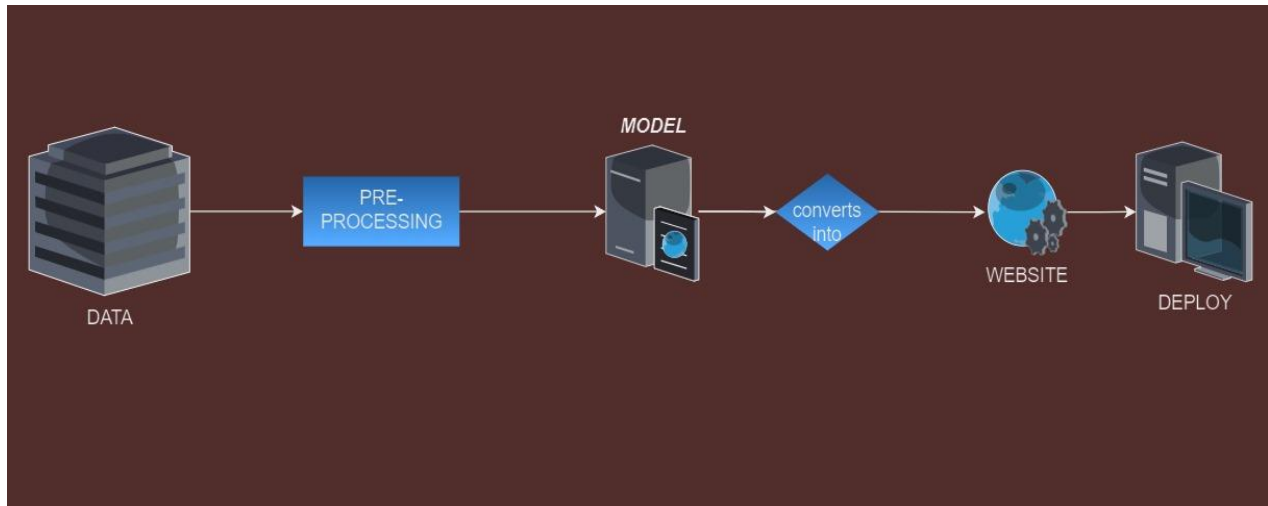
- ➢ Data Import: The first step involves importing the movie dataset from Kaggle.
- ➢ Data Preprocessing: We will preprocess the data by handling missing values, cleaning text data, and extracting relevant features like genres, keywords, and director.
- ➢ Vectorization: Text data will be converted into numerical vectors.
- ➢ Cosine Similarity: Instead of using Euclidean distance, we will calculate the similarity between movies using cosine distance. This helps in capturing the semantic similarity between movie content.
- ➢ Stemming: We will apply stemming to reduce words to their root form, improving the effectiveness of text processing and similarity calculations.
- ➢ Model Building: We will build a content-based recommendation model using preprocessed data and cosine similarity calculations.
- ➢ Website Conversion: The model will be integrated into a web application where users can input a movie and receive recommendations.

➢ Deployment: Finally, the recommender system will be deployed to a web server for real-time usage.

5. Conclusion:

This project aims to provide a robust content-based movie recommendation system that leverages advanced natural language processing techniques. By focusing on content similarity and utilizing cosine distance calculations, we expect to deliver accurate and personalized movie recommendations to users.

6. Future works:

In the future, after deploying the initial version of the content-based movie recommender system and gathering user feedback, we can explore several avenues for enhancing and refining the system.