

### **Project Synopsis**

Name	ARSHA P JOY
USN	221VMTR01438
Elective	Artificial Intelligence
Date of Submission	

# **CONTENT BASED MOVIE RECOMMENDATION SYSTEM**

## **Executive Summary:**

In the world of enhanced entertainment possibilities, navigating through the sea of movies can be overwhelming and time-consuming for viewers. This project aims to develop a content-based movie recommender system that will utilize the data from Rotten Tomatoes. By analysing various movie features such as audience scores, critic reviews, genres, directors, ratings, and more, the system will offer personalized recommendations tailored to individual preferences. This project seeks to streamline the movie-finding process, enhance user experience, promote diversity in movie consumption, and enhance the audience range widely by exploring even unheard or underrated areas.

## **Objectives and Scope:**

To build a content-based movie recommender system utilizing machine learning algorithms.

To enhance user experience by providing personalized movie recommendations based on individual preferences.

To promote diversity in movie consumption by exposing users to a broader range of films, including hidden gems and lesser-known titles.

## **Methodology:**

Developing a content-based movie recommendation system utilizing Rotten Tomatoes data involves a systematic approach encompassing data preprocessing, feature extraction, model selection, and evaluation. Below is a detailed methodology outlining each step:

### **Data Preprocessing:**

Begin by loading the Rotten Tomatoes movie datasets (rotten\_tomatoes\_movies.csv and rotten\_tomatoes\_movie\_reviews.csv) into a suitable data structure and conduct initial data cleaning tasks such as handling missing values, removing duplicates, and ensuring data integrity. Convert relevant features into appropriate data types and standardize formats for consistency.

Perform text preprocessing on textual features such as movie titles, genres, directors, and writer names. This includes tokenization, lowercasing, and removing punctuation.

#### Feature Extraction:

Extract features from the pre-processed data that are relevant for content-based recommendation. Features to consider may include movie genres, director names, writer names, audience scores, critic reviews, and release dates.

Utilize techniques such as one-hot encoding for categorical features like genres and directors. Extract numerical features such as audience scores and runtime, scaling them if necessary to ensure uniformity.

#### Model Selection:

Choose appropriate algorithms for content-based recommendation, considering the nature of the problem and available data. Commonly used techniques include cosine similarity, TF-IDF (Term Frequency-Inverse Document Frequency), and vector embeddings. Implement algorithms that calculate similarity scores between movies based on their features. Explore variations of content-based filtering algorithms, such as item-item or user-item similarity.

#### Model Training and Recommendation Generation:

Train the selected content-based recommendation model using the pre-processed data. Implement recommendation generation logic that takes a user's preferences or previously liked movies as input. Calculate similarity scores between the user's preferences and all movies in the dataset using the trained model. Rank movies based on their similarity scores and recommend the top-N most similar movies to the user.

#### Evaluation:

Evaluate the performance of the content-based recommendation system using appropriate metrics. Metrics to consider may include precision, recall, F1-score, and Mean Average Precision (MAP). Additionally, conduct the qualitative evaluation by assessing the relevance and diversity of recommended movies. Validate the effectiveness of the recommendation system through user feedback and user testing, if possible.

By following this methodology, the development of a robust content-based movie recommendation system that leverages Rotten Tomatoes data to provide personalized and relevant movie recommendations to users will be made possible.

### Preliminary Findings & Expected Results

Existing movie recommender systems often rely on collaborative filtering or hybrid approaches to generate recommendations. These models typically analyse user ratings and movie metadata to make predictions. However, these approaches may not fully capture the essence of individual preferences or account for diverse tastes. By leveraging Rotten Tomatoes data, including audience scores, critic reviews, and movie features, the proposed recommender system is expected to offer more accurate and personalized recommendations. The expected results include improved user satisfaction, enhanced movie discovery, and increased diversity in movie consumption.