PHASE-2

**DELIVERING PERSONALIZED MOVIE RECOMMENDATION WITH AN AI-DRIVEN MATCHMAKING SY STEM**

**Student Name : R.Senthamarai**

**Register Number : 510523106037**

**Institution : Bharathidasan Engineering college**

**Department : Electronics and**

**Communication department**

**Date of submission :08-05-2025**

**1. Problem Statement**

In Phase I, we established the foundation for a content-based movie recommendation system. The primary challenge was to match user preferences with the most suitable content by analyzing metadata, genres, user behavior, and ratings.

**2. Project Objectives**

**Phase I Objectives**:

\* To create a basic recommendation model using content filtering.

\* To process and clean raw movie and user rating data.

\* To perform exploratory data analysis (EDA) for insights into viewing behavior.

**Phase II Enhancements**:

1. Introduce user-to-user matchmaking based on preferences.
2. Incorporate AI-driven models for enhanced recommendation accuracy.
3. Build an interactive system capable of real-time suggestions.

**3. Flowchart of the Project Workflow**

Load Dataset -> Data Preprocessing -> EDA -> Feature Engineering -> Model Building -> Personalized Recommendations -> Result Visualization & Insights



**4. Data Description**

**Dataset:**

We use a modified version of the MovieLens dataset (movies.csv, ratings.csv, and users.csv).

1. movies.csv: movieID, title, and genres.
2. ratings.csv: userID, movieID, rating, and timestamp.
3. users.csv: demographic information.

**5. Data Preprocessing**

**Steps involved:**

1. Handled missing values.
2. Multi-hot encoded genres.
3. Normalized rating scales.
4. Merged datasets for training.

**6. Exploratory Data Analysis (EDA)**

**Key Findings:**

1. Action and Drama are the most watched genres.
2. Ratings mostly 3.5–4.0 stars.
3. Viewing preferences vary by demographics.

**7. Feature Engineering**

1. Created user profiles from genre preferences.
2. Used TF-IDF for genre/title vectors.
3. Built similarity matrix with cosine similarity.

**8. Model Building**

**Models Used:**

1. -Collaborative Filtering (KNN)
2. Hybrid Model (combined approach)
3. content-based filtering

**9. Visualization of Results & Model Insights**

1. Plotted model metrics (Precision, Recall, F1)
2. Visualized user clusters and preferences.

**10. Tools and Technologies Used**

1. Python (Pandas, NumPy, Scikit-learn)
2. Jupyter Notebook
3. Matplotlib & Seaborn
4. GitHub
5. Google Colab
6. Streamlit (optional UI)

**11. Team Members and Contributions**

R.Senthamarai - Data Analyst , Preprocessing & EDA .

U. Kaviya - ML Engineer , Model Building .

V. Shalini - Visualization Expert , Dashboard & Results .

R. Ramya - Documentation , Report & GitHub .