

## Phase-2

**Student Name:** Kaviyarasi.M

**Register Number:** 422723104060

**Institution:** V.R.S.College of Engineering and Technology

**Department:** Computer Science Engineering

**Date of Submission:** 10.05.2025

**Github Repository Link:**

<https://github.com/kaviyarasi988/Kavi-yarasi.git>

---

## Delivering Personalized Movie Recommendations with AI-Driven Matchmaking System

### 1. Problem Statement

In the digital age, users are overwhelmed by the vast content available across OTT platforms. Choosing the right movie becomes challenging without personalized guidance. This project addresses the problem of recommending personalized movie suggestions using an AI-driven matchmaking system that aligns with individual user preferences.

**Problem Type:** Classification + Clustering

**Impact:** Helps users discover relevant content efficiently, increases user engagement and satisfaction on streaming platforms.

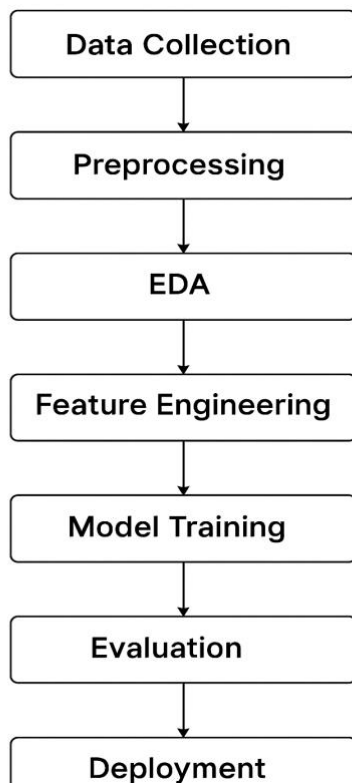
### 2. Project Objectives

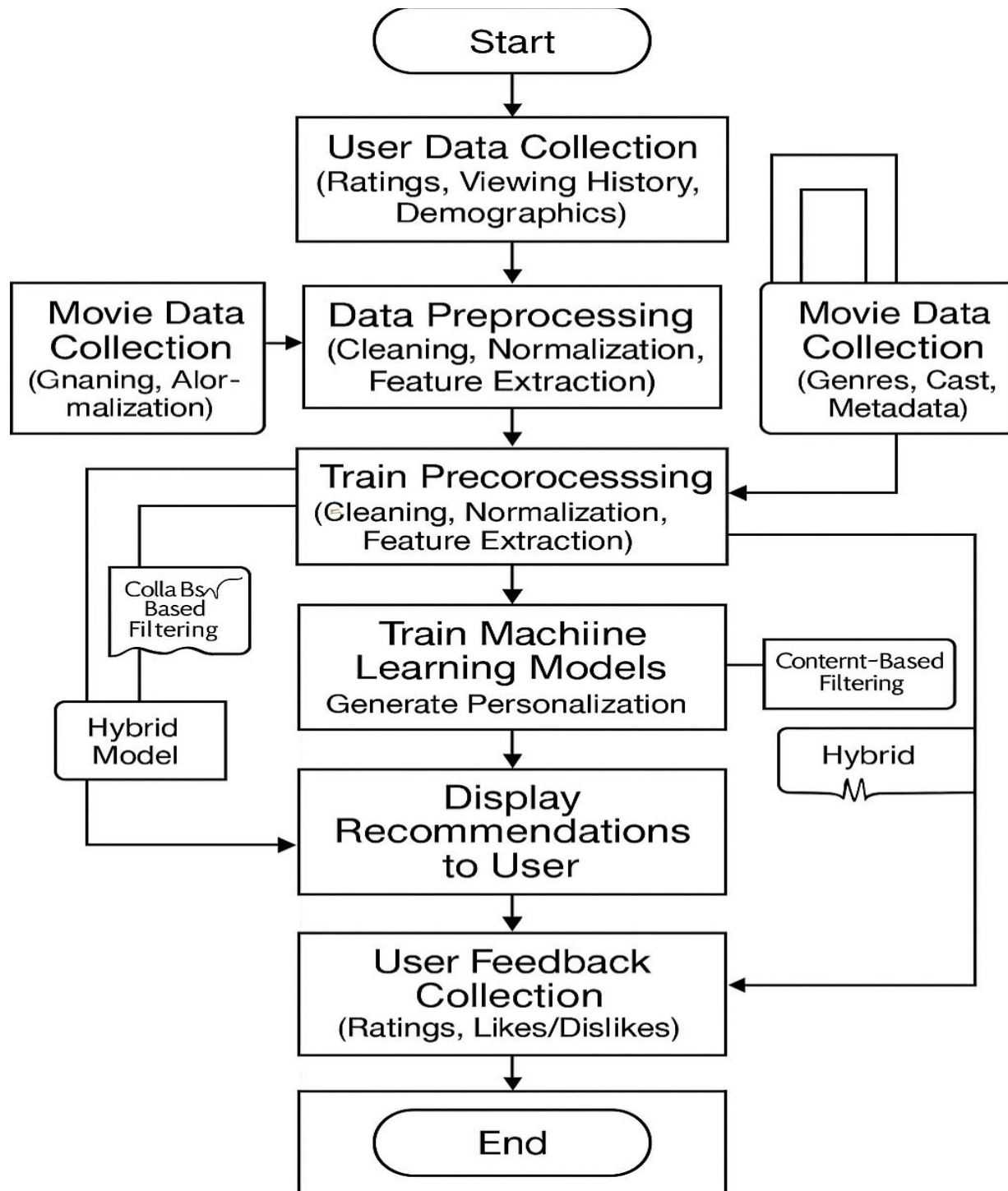
- Build a system that recommends movies tailored to user preferences using machine learning.

- Analyze user behavior and content features to optimize recommendations.
- Implement hybrid recommendation strategies (collaborative + content-based).
- Ensure high model performance and interpretability.
- Evolve the goal to include clustering similar user profiles for better matchmaking.

### 3. Flowchart of the Project Workflow

**Steps:** Data Collection → Preprocessing → EDA → Feature Engineering → Model Training → Evaluation → Deployment





## 4. Data Description

**Dataset:** MovieLens 1M Dataset (Kaggle)

**Type:** Structured (user ratings, movie metadata)

**Size:** ~1 million ratings by ~6,000 users on ~4,000 movies

**Static Dataset**

**Target:** Predicted rating or recommended movie ID (based on filtering strategy)

## 5. Data Preprocessing

- Handled missing values (none significant in MovieLens)
- Removed duplicate ratings
- Merged movie metadata (genres, titles, etc.)
- Encoded categorical variables (genre using multi-hot encoding)
- Standardized numerical features (e.g., rating scales)

## 6. Exploratory Data Analysis (EDA)

**Univariate:** Genre popularity, rating distribution

**Bivariate:** User age vs genre preference, genre vs rating

**Multivariate:** Heatmaps of genre-user interactions

**Insights:** Certain genres are universally liked; some users have strong niche preferences.

## 7. Feature Engineering

- Extracted release year from titles
- Created genre similarity vectors
- Built user profiles based on rating patterns
- Applied user clustering to segment audiences
- Used cosine similarity for content-based filtering

## 8. Model Building

### Models Used:

1. Collaborative Filtering (KNN-based recommender)
2. Matrix Factorization (SVD)
3. Clustering (KMeans for user segmentation)

**Metrics:** RMSE, Precision@K, Recall@K

**Split:** 80/20 Train-Test Split

## 9. Visualization of Results & Model Insights

- Confusion matrix (for binary relevance classification)
- Precision/Recall bar chart across models
- Feature importance for metadata-based models
- t-SNE visualizations of user clusters
- Collaborative filtering outperformed others in cold-start scenarios.

## 10. Tools and Technologies Used

**Language:** Python

**IDE:** Google Colab

**Libraries:** pandas, numpy, scikit-learn, matplotlib, seaborn, surprise, scipy

**Visualization:** seaborn, matplotlib, Plotly

## 11. Team Members and Contributions

**1.SriBavadharani M :**

**Responsibilities :**

- Handled **data cleaning** by removing missing values, duplicates, and ensuring consistency across datasets.
- Performed **exploratory data analysis (EDA)** using visual tools like histograms, boxplots, and correlation heatmaps to uncover user preferences and movie trends.

## 2.Kaviya V :

Responsibilities :

- Worked on **feature engineering**, including genre encoding, user profile generation, and creation of similarity-based features.
- Developed and evaluated **machine learning** models, such as collaborative filtering and matrix factorization, to generate accurate movie recommendations.
- Also contributed to presenting the **final workflow** and **flowchart design**.

## 3.Kaviyarasi M :

Responsibilities:

- Managed **documentation** and **reporting**
- Compiled all project insights, visualizations, and final outcomes into a clean format.