Phase-2

Student Name: Vishanth V

Register Number: 410723106036

Institution: Dhanalakshmi College Of Engineering

Department: Electronics and Communication Engineering

Date of Submission: 07/05/2025

Github Repository Link:

https://github.com/Vish2327/Nm_Vishanth

---

1. Problem Statement

Delivering Personalized Movie Recommendations with an AI-Driven

With the exponential growth of digital content, users often struggle to find movies that match their personal preferences. Traditional recommendation systems (e.g., popularity-based or genre-based) fail to capture user-specific tastes, leading to poor user engagement. Our project aims to design a personalized, AI-driven movie recommendation system that leverages user behavior and preferences.

Problem Type: Supervised learning (Classification/Regression) or Unsupervised learning (Clustering), depending on approach.

Relevance: Enhancing user experience in streaming platforms by offering customized movie suggestions, improving retention, and engagement.

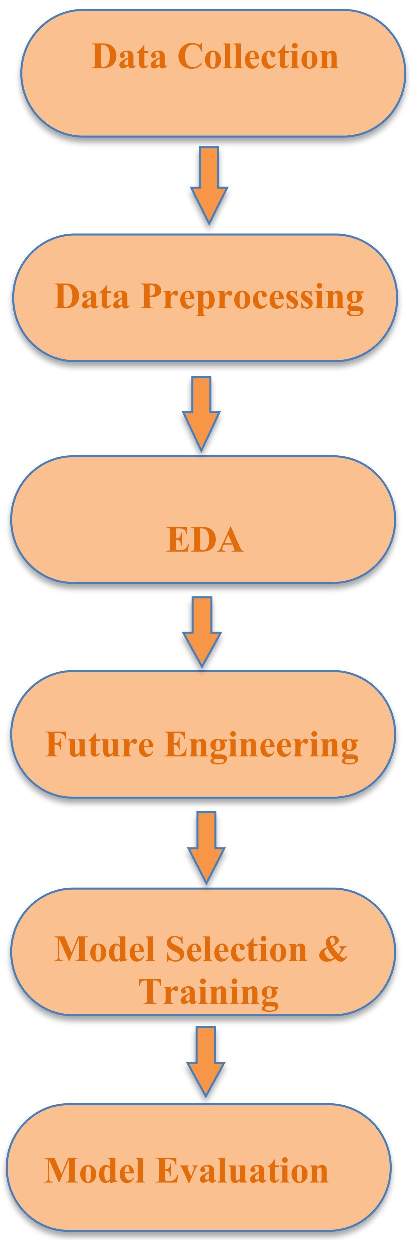
2. Project Objectives

Technical Goals:

- Analyze user and movie data to extract meaningful patterns.
- Develop a recommendation engine using collaborative and/or content-based filtering.
- Achieve high recommendation accuracy (e.g., RMSE below a threshold).
- Ensure scalability and user-specific relevance.

Updated Goals: After data exploration, the focus expanded to hybrid models combining collaborative and content-based filtering for better personalization.

3. Flowchart of the Project Workflow



4. Data Description

- **Dataset:** MovieLens dataset (from GroupLens / Kaggle)
- **Type:** Structured data
- **Records & Features:** ~100,000+ ratings; includes userId, movieId, rating, timestamp, genres, etc.
- **Nature:** Static
- **Target Variable:** Rating (for supervised model) or User-Movie interaction matrix (for unsupervised collaborative filtering)

5. Data Preprocessing

- Handled missing values and removed duplicates.
- Converted timestamps to datetime format.
- One-hot encoded genres and encoded user/movie IDs.
- Normalized rating values where necessary.
- Ensured consistency across merged datasets (e.g., movies and ratings).

6. Exploratory Data Analysis (EDA)

Univariate Analysis: Histograms for rating distributions, movie popularity, genre frequency.

Bivariate Analysis: Correlation between user rating behavior and movie features.

Insights:

- Users tend to rate popular movies more frequently.
- Certain genres like Drama and Comedy dominate the dataset.

7. Feature Engineering

- Created user-movie interaction matrix.
- Extracted movie release year from title.
- Binned rating values to categorize user sentiment.
- Engineered features like average user rating, genre similarity scores.
- Considered dimensionality reduction (e.g., SVD for matrix factorization).

8. Model Building

Models Used:

- Collaborative Filtering (Matrix Factorization via SVD)
- Content-Based Filtering (TF-IDF on movie descriptions)

Why Chosen: Well-suited for recommendation tasks; SVD handles sparse data efficiently.

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

9. Visualization of Results & Model Insights

- Plotted RMSE for different models.
- Displayed confusion matrix for classification of rating levels (if applicable).
- Featured bar chart showing most influential features (genres, average ratings).
- ROC curves for binary classification variants.

10. Tools and Technologies Used

Language: Python

IDE: Google Colab

Libraries: pandas, numpy, matplotlib, seaborn, scikit-learn, surprise, lightFM, plotly

11. Team Members and Contributions

NAME	ROLE	RESPONSIBILITY
Vishanth V	Leader	EDA, Feature Engineering, Model Building, Documentation
Vinoth V	Member	Data Cleaning, Visualization, Reporting
Santhkumar C	Member	Model Evaluation, Deployment Setup