Phase-1 Submission Template

Student Name: Darsini B

Register Number: 731123104010

Institution: Government College Of Engineering, Erode.

Department: Computer Science and Engineering.

Date of Submission: 28-04-2025

1.Problem Statement

With an overwhelming amount of movie content available across streaming platforms, users often face difficulty in selecting movies they will enjoy. Traditional recommendation systems rely heavily on simple genre or popularity filters, leading to less personalized experiences. This project aims to solve this problem by developing an AI-driven matchmaking system that provides highly personalized movie recommendations based on individual user preferences, behaviors, and tastes.

2.Objectives of the Project

- To build a recommendation system that predicts and suggests movies personalized to user interests.
- To analyze user behavior and movie metadata to improve recommendation accuracy.
- To implement a matchmaking system that dynamically adapts to user feedback and evolving tastes.

3.Scope of the Project

Features: Movie ratings, genres, directors, actors, user reviews, watch history.

Constraints:

- Dataset limited to publicly available data.
- Focus on supervised machine learning models and deep learning-based recommender systems.

- Initial deployment as a simple web application.

4.Data Sources

- Dataset: MovieLens dataset from Kaggle (public dataset).
- Source: Kaggle.com.
- Type: Static dataset.
- Additional synthetic user interaction data might be generated for training.

5.High-Level Methodology

1. **Data Collection:**
   - Download MovieLens dataset from Kaggle.
2. **Data Cleaning:**
   - Handle missing values, remove duplicates, ensure consistent data formatting.
3. **Exploratory Data Analysis (EDA):**
   - Use histograms, correlation matrices, and clustering to understand user behavior and movie trends.
4. **Feature Engineering:**
   - Generate new features like user genre preference scores, movie popularity ranks.
5. **Model Building:**
   - Based Filtering (based on movie metadata).
   - Collaborative Filtering (user-user and item-item).

- Hybrid models (combining collaborative and content-based approaches).
- Deep Learning models (e.g., Neural Collaborative Filtering).

6. Model Evaluation:

- Metrics: RMSE, Precision@K, Recall@K, F1-score.
- Cross-validation using train-test splits.

7. Visualization & Interpretation:

- Interactive dashboards using Plotly/Seaborn for showing insights and recommendation examples.

8.Deployment:

- A basic web application using Streamlit or Flask to allow users to input preferences and receive recommendations.

6. Tools and Technologies

Programming Language:

- Python.

Notebook/IDE:

- Jupyter Notebook, Google Colab.

Libraries:

- Data Processing: pandas, numpy. Visualization:
- matplotlib, seaborn, plotly.
- Modeling: scikit-learn, surprise, TensorFlow/Keras.
- Deployment: Streamlit or Flask.

Optional Tools for Deployment:

•Streamlit for easy frontend development.

7.Team Members

- Member 1: Darsini B -Data Collection and Cleaning.
- Member 2: Yuthika M S -Exploratory Data Analysis.
- Member 3: Sambavi A -Model Building and Evaluation.
- Member 4: Kaviya S -Visualization and Dashboard Development.
- Member 5: Bharani E K -Deployment and Final Presentation.