Phase-1 Submission

Student Name: Manjima M.

Register Number: 712523205038

Institution: PPG institute of technology

Department: B tech Information Technology

Date of Submission: 09.05.2025

1. Problem Statement

With the rapid expansion of digital content platforms, users often struggle to discover movies that match their personal preferences. Traditional recommendation systems fall short by relying heavily on basic metrics such as genre or user ratings. This project aims to solve this issue by creating a personalized movie recommendation system using AI-driven matchmaking, considering factors like user behavior, viewing history, preferences, and content features to enhance user satisfaction and engagement.

2. Objectives of the Project

- *To build a personalized movie recommendation system using machine learning and deep learning techniques.*
- *To analyze user preferences and movie features to generate highly tailored suggestions.*
- *To improve recommendation accuracy by incorporating collaborative and content-based filtering.*
- *To provide real-time recommendations through a user-friendly interface or app.*

3. Scope of the Project

Features:

- *User profiling and behavior analysis.*
- *Content-based and collaborative filtering recommendation models.*
- *Real-time or near real-time recommendation output.*

Limitations:

- *Limited to available datasets (e.g., IMDb, MovieLens).*
- *Deployment scope may be limited to a demo web app or dashboard.*

4. Data Sources

- ***MovieLens Dataset*** *:* https://www.kaggle.com/datasets/justsahil/movielens-32m
- ***IMDb API or Kaggle Dataset*** *:* https://www.mdpi.com/2076-3417/11/20/9381
- ***Synthetic data*** *: https://www.kaggle.com/datasets/tmdb/tmdb-movie-metadata*

5. High-Level Methodology

Data Collection*: Download datasets from Kaggle and MovieLens; extract movie details from IMDb.*

Data Cleaning*: Remove duplicates, handle missing values, standardize movie titles and formats.*

Exploratory Data Analysis (EDA)*: Use visualizations (bar charts, word clouds, correlation plots) to uncover insights.*

Feature Engineering*: Create features like user genre preference vectors, tag embeddings, movie similarity matrices.*

Model Building*:*

Collaborative filtering using matrix factorization or neural networks.

Content-based models using cosine similarity and NLP for movie descriptions.

Hybrid models combining both.

Model Evaluation*: Use metrics like RMSE, precision@k, recall@k, and diversity score.*

***Visualization & Interpretation**: Use dashboards to display recommended movies, user profiles, and model insights.*

***Deployment**: Flask app to demonstrate real-time recommendations.*

6. Tools and Technologies

- ***Programming Language**: Python*
- ***Notebook/IDE**: Google Colab or Jupyter Notebook*
- ***Libraries**: pandas, numpy, seaborn, matplotlib, scikit-learn, Surprise, TensorFlow/Keras, nltk or spaCy*
- ***Optional Tools for Deployment**: Streamlit, Flask, Gradio.*

7. Team Members and Roles

S NO	NAME	ROLE	DESCRIPTION
1.	**Kaviya Bharathi K.**	Data Collection, EDA	Movie and user data will be gathered from MovieLens, IMDb, and other open sources.
2.	Manjima M.	Evaluation, Feature Engineering	Model accuracy will be measured using metrics like RMSE, precision@k, and recall@k.
3.	Dhinesh Kumar B.	Model Building, Visualization	Key insights and recommendations will be displayed using interactive charts and graphs.
4.	Vishnu Kumar K.	Deployment	The model will be deployed as a web app using tools like Streamlit or Flask for real-time recommendations.
5.	Balaji G.	UI/UX, Final Report Preparation	A detailed report will summarize the project workflow, results, and key findings.