**Phase-1 Submission**

**Student Name:** Sai Mouleeshwar S.M.

**Register Number:** 712523106015

**Institution:** PPG institute of technology

**Department:** Electronics And Communication Engineering

**Date of Submission:** 28/04/2025

1. **Problem Statement**

*This project aims to solve the problem by developing a personalized recommendation system that accurately matches users with movies they are likely to enjoy and helps them to be connected without any confusions in selecting movies.*

1. **Objectives of the Project**

*By the end of this project, we aim to achieve the following:*

* *Develop a robust AI model capable of predicting user movie preferences based on historical viewing data and other relevant factors.*
* *Create a personalized movie recommendation system that suggests a ranked list of movies tailored to individual users.*
* *Gain insights into the key factors that influence user movie choices and identify patterns in viewing behavior.*

1. **Scope of the Project**

*This project will focus on:*

* *Analyzing user-movie interaction data to understand viewing patterns and preferences.*
* *Evaluating the performance of the chosen recommendation algorithms. Developing a basic prototype to showcase the recommendation functionality.*

1. ***Data Sources***

*We will utilize the Movie Lens dataset (e.g. Movie Lens 25M Dataset), which is a publicly available dataset from Group Lens.*

*Source: Group Lens (available on their website and platforms like Kaggle).*

*Nature: Static*

*Data Source link:*

***Movie Lens Dataset****:* [***https://www.kaggle.com/datasets/dnyaneshyeole/10000-most-popular-english-movies-2023***](https://www.kaggle.com/datasets/dnyaneshyeole/10000-most-popular-english-movies-2023)

***IMDb Dataset (Kaggle*):** [***https://www.kaggle.com/datasets/lakshmi25npathi/imdb-dataset-of-50k-movie-reviews***](https://www.kaggle.com/datasets/lakshmi25npathi/imdb-dataset-of-50k-movie-reviews) 1

1. ***High-Level Methodology***

***Data Collection*** *– Download the chosen Movie Lens dataset (e.g., the 25M version) from its source.*

***Data Cleaning*** *– Inspect the dataset for missing values in key columns (e.g., user ID, rating). We will decide on a strategy to handle missing values.*

***Exploratory Data Analysis (EDA)*** *– Visualize the distribution of user ratings to understand the general sentiment. Analyze the number of ratings per user and per movie to identify popular movies and active users.*

***Feature Engineering*** *– Create user-item interaction matrices, which are fundamental for collaborative filtering. If incorporating content-based filtering, we might preprocess the movie genre information (e.g., one-hot encoding).*

***Model Building*** *– Collaborative Filtering: Implement memory-based collaborative filtering techniques such as user-based and item-based k-Nearest Neighbours (k-NN) using similarity metrics like cosine similarity or Pearson correlation.*

***Model Evaluation*** *– split the dataset into training and testing sets to evaluate the model's ability to generalize to unseen data. Precision@k, Recall@k, and F1-score@k to evaluate the relevance of the top-k recommended movies.*

***Visualization & Interpretation*** *-Visualize the performance of different models using bar charts or tables to compare their evaluation metrics. Interpret the learned latent factors from matrix factorization.*

***Deployment*** *– For this project, the "deployment" will likely involve creating a Jupyter Notebook or a simple Python script that takes a user ID as input and outputs a list of recommended movies.*

***6. Tools and Technologies***

* ***Programming Language*** *– Python is used.*
* ***Notebook/IDE*** *– Jupyter Notebook or Google Colab*
* ***Libraries*** *–pandas: For data manipulation and analysis numpy: For numerical computations. seaborn and matplotlib : For data visualization. scikit-learn: For model building.*
* ***Optional Tools for Deployment****: Flask, Stream lit.*

***7. Team Members and Roles:***

|  |  |  |  |
| --- | --- | --- | --- |
| ***S NO*** | ***TEAM MEMBERS*** | ***ROLE*** | ***DESCRIPTION*** |
| ***01.*** | ***Sai Mouleeshwar S.M.*** | *Project Manager and Documentation Lead.* | *A Project Manager oversees project execution and team coordination, while a Documentation Lead ensures accurate, organized, and accessible project documentation.* |
| *02.* | *Selvam A.* | *Data Analyst.* | *A Data Analyst collects, processes, and interprets data to help organizations make informed business decisions.* |
| *03.* | *Madhan Raj R.* | *Visualisation and Deployment Specialist.* | *It is the graphical representation of data and information to make insights easier to understand and analyse.* |
| *04.* | *Vignesh V.* | *Machine Learning Engineer and Algorithm Developer.* | *It designs, builds, and deploys algorithms and models that enable machines to learn from data and make predictions or decisions.* |