**Phase-1 Submission Template**

**Student Name:** Kishordharan P

**Register Number:** 622023104049

**Institution:** Paavai College of Engineering

**Department:** BE - Computer Science & Engineering

**Date of Submission:** 24-04-2025

**1.Problem Statement**

*Cinemaphiles are all around the world, and most of them are often confused about what to watch next, and spent huge amount of time, literally hours not realizing this isn't what they were looking for.*

**2.Objectives of the Project**

*A personal AI which suggest movies based on previous watch history, and which will learn, and suggest movies based on users previous likings, such as the story, the cast, or the cinematography of the movie.*

**3.Scope of the Project**

*It should be able to keep track of movies which the user watched without any time constraint and also the user's implicit or explicit review, such as whether they skipped any part of the movie or what they share. The major problem we're going to expect is the sheer amount of data which needs to be handled and requirement of heavy computing power.*

**4.Data Sources**

*Using data set isn't ideal for movie recommendation, because of the vast amount of movies that will be released in a small time interval we can simply state its very fluctuating. so using an API is the best choice for its regular updates. public API such as IMDB's API and other's in the market which are free to use will be the efficient idea.*

**5.High-Level Methodology**

* **Data Collection** – *IMDB's API or other free to use API.*
* **Data Cleaning** – *There is no need for any major cleansing, The data we get from these programs are well cleaned and up to date.*
* **Exploratory Data Analysis (EDA)** – *We are planning to use vectors one defines the users preference and n vectors for each n movies we are going to compare*
* **Feature Engineering** – *These mathematical algorithms have been existed for a while we use them for a more personalized version which removes bias unlike any similar programs.*
* **Model Building** – *Single value decomposition (SVD), Alternating Least Squares (ALS)..*
* **Model Evaluation** – *Uses a simple validation strategy by collecting the data's on how much the user likes the movies based on the suggestion in a huge scale, where N>=100.*
* **Visualization & Interpretation** – *The finding's will be displayed in a form of a ranked list, where the most likely vectored movies based on the users taste are listed in an ascending order.*
* **Deployment** – Our team haven't planned on releasing this project to the public yet.

**6.Tools and Technologies**

* **Programming Language** – *Python for it's huge collection of libraries.*
* **Notebook/IDE** – *Google collab for cloud computing.*
* **Libraries** –  *pandas, numpy, scikit-learn for cosine similarity and vector normalizations and fastapi for api's.*

**7.Team Members and Roles**

Jaya Surya M - Team Leader

Kishordharan P - Programming & Designing

Asbel Jino K - Graphical Designing and Testing

Jeevi Kannan A - Functionality & Efficiency Testing