Movie Recommendation System

Zehra Yıldız Açıkgül, *CENG489*

I. THE DATA SET

FOR my movie recommendation system project, i will select a comprehensive dataset that includes information about movies, such as genre, cast, crew, ratings, reviews, and user preferences. I aim to gather data from reliable sources like IMDb, TMDb, or Kaggle datasets.

II. THE PROBLEM

The problem i aim to solve is the overwhelming amount of content available on streaming platforms, making it challenging for users to discover movies that match their preferences. This leads to user dissatisfaction and decreased engagement with the platform. By implementing a movie recommendation system, i aim to enhance user experience by providing personalized recommendations tailored to individual preferences, ultimately increasing user satisfaction and platform engagement.

III. METHODOLOGY

Machine Learning Algorithms for Movie Recommendation Systems

In movie recommendation systems, various machine learning algorithms play a crucial role in filtering information and conducting data mining to achieve desired outcomes. Understanding the functioning of these algorithms is vital for selecting the most suitable one for a specific task within recommender systems.

K-Means Clustering

K-Means clustering is a simple collaborative filtering approach that categorizes users based on their interests. It groups users with similar interests together to make recommendations. This algorithm utilizes common user features such as age, gender, movie preferences, and viewing history to group users into clusters that represent their characteristics.

Measurement of Similarities

The first step involves finding similarities in user features between new users and existing ones in the system. Features such as age, viewing history, and geographic locations are considered. Similarities can be calculated using metrics like Pearson correlation or cosine similarity, either item-based or user-based.

Selection of Neighbors

After measuring similarities, the algorithm selects neighbors with similar preferences to the new user. This step involves identifying users with comparable characteristics and viewing habits.

Prediction Computation

Predictions are computed based on the closest neighbors found in the system database. The algorithm suggests movies that are popular among similar users, utilizing the closest neighbors' preferences and viewing patterns.

By leveraging K-Means clustering and its steps of similarity measurement, neighbor selection, and prediction computation, movie recommendation systems can provide personalized suggestions tailored to individual users' preferences and characteristics.

IV. THE EVALUATION METHOD

I will evaluate the performance of our recommendation system using standard evaluation metrics such as precision, recall, F1-score. Additionally, we i conduct user surveys and A/B testing to gather feedback and measure user satisfaction with the recommended movies.

V. THE TIME PLAN

1) Due Mar. 26: Project topic
2) Due Apr. 02: Project Proposal
3) Due May. 07: Intermediate Report
4) Due May. 28 – June. 04: Demo, presentation, submission of all codes
5) Demo/presentation: Jan. 01, during class hours
6) Due Jan. 04: Final Report

VI. BACKUP PLAN

In case of unexpected challenges or data limitations, i will explore alternative datasets or simplify the recommendation system's scope to ensure timely completion of the project. Additionally, i will prioritize key features and functionalities to deliver a functional prototype even if certain components cannot be fully implemented.