Movie Recommendation System Report

# 1. Introduction

This report presents the development of a Movie Recommendation System using the MovieLens dataset. The project explores collaborative filtering techniques to recommend movies based on either user similarity or item similarity. The system provides personalized recommendations and evaluates their accuracy using Precision@K metrics.

# 2. Methodology

## 2.1 Dataset

The dataset was obtained from KaggleHub and consists of three files:

* movie\_info.csv → Movie titles, genres, and IDs.
* ratings.csv → User ratings for movies.
* user\_demographics.csv → Age, gender, and other attributes of users.

## 2.2 Preprocessing

The ratings and movie information were merged to form a comprehensive dataset. A train-test split was performed per user, ensuring each user's ratings were divided into 80% training and 20% testing sets. Users with fewer than 5 ratings were kept only in the training set.

## 2.3 Collaborative Filtering Approaches

Two collaborative filtering approaches were implemented:

* User-Based Collaborative Filtering: Identifies similar users using cosine similarity and recommends movies based on neighbors' preferences.
* Item-Based Collaborative Filtering: Identifies movies similar to those already rated by the user and generates recommendations.

# 3. Evaluation

The evaluation metric used was Precision@K, which measures the proportion of recommended movies that are relevant (i.e., present in the test set). Both user-based and item-based collaborative filtering were assessed using this metric.

# 4. Visualization

The system included visualizations to provide insights into the dataset, such as:

* Distribution of Movie Ratings
* Top 10 Most Frequently Rated Movies

# 5. Conclusion

The Movie Recommendation System successfully implemented both User-Based and Item-Based Collaborative Filtering. It demonstrated the ability to provide personalized recommendations and achieved reliable performance as evaluated by Precision@K. Future improvements could include integrating hybrid approaches and leveraging deep learning models for enhanced accuracy.