Movie Recommendation System Report

Team Members: Yuvraj and Shivansh

1. Introduction

The goal of this project is to build a **movie recommendation system** using **collaborative filtering and content-based filtering techniques**. We analyze a dataset containing movies, ratings, and tags to recommend movies based on user preferences and movie characteristics.

2. Data Collection & Preprocessing

We used three datasets:

- **movies.csv:** Contains movie IDs, titles, and genres.
- **ratings.csv:** Contains user ratings for different movies.
- **tags.csv:** Contains user-assigned tags for movies.

Data Cleaning:

- Checked for **missing values** and removed them.
- Extracted **movie release years** from titles.
- Merged datasets to form a comprehensive dataset for analysis.

3. Exploratory Data Analysis (EDA)

- **Genre Analysis:** Counted occurrences of genres to identify popular ones.
- **Rating Distribution:** Analyzed the range of ratings given by users.
- **Popular Movies:** Identified movies with high ratings and a significant number of reviews.

- **Word Cloud Visualization:** Generated a word cloud for genre frequency.
- **Genre Popularity:** Created a histogram to visualize genre distribution.

4. Recommendation Models

A. Collaborative Filtering (SVD)

We applied **Singular Value Decomposition (SVD)** to the **user-item utility matrix**, which predicts ratings by factorizing the matrix into **latent features**.

Evaluation Metrics for SVD:

- **Mean Squared Error (MSE):** Measures prediction error.
- **Mean Absolute Error (MAE):** Measures the average absolute prediction error.
- **R^2 Score:** Indicates how well the model fits the data.
- **Classification Metrics:** We classified ratings as "Liked" (≥3.5) or "Not Liked" and evaluated:
  - **Accuracy, Precision, Recall, F1-Score, AUC-ROC**
  - **Confusion Matrix visualization**

B. Hybrid Model (SVD + Bayesian Average)

To improve recommendation accuracy, we incorporated **Bayesian Averaging** of movie ratings along with SVD predictions. The hybrid model was constructed as:

$$\text{Hybrid Prediction} = \alpha \times \text{SVD Prediction} + (1 - \alpha) \times \text{Bayesian Average}$$

where **$\alpha = 0.5$** (tunable parameter).

Hybrid Model Performance Comparison:

Model	MSE ↓	MAE ↓	R^2 Score ↑
SVD	2.0893	1.1606	-0.9224
Hybrid	0.6042	0.6288	0.4440

The **hybrid model outperformed** SVD alone by incorporating Bayesian averaging, making recommendations more stable and reliable.

5. Conclusion & Future Scope

- **Our recommendation system effectively suggests movies using collaborative filtering and hybrid approaches.**
- The **hybrid model** improves accuracy by considering both user preferences and overall movie popularity.
- **Future Improvements:**
  - Implement **deep learning**based recommendations.
  - Use **genre embeddings** for content-based recommendations.
  - Deploy the model as a **web-based recommendation system**
