1. Introduction

Recommendation systems play a crucial role in enhancing user experience in various domains, such as entertainment, e-commerce, and online learning. This project focuses on building a movie recommendation system using multiple approaches, including **content-based filtering, collaborative filtering (user-based and item-based), matrix factorization (SVD), and Neural Collaborative Filtering (NCF)**. The objective is to compare these methods and evaluate their effectiveness in predicting user preferences.

2. Methodology

Data Preprocessing

The **MovieLens dataset** was used, containing movie ratings by different users. The preprocessing steps included:

- **Handling missing data** by filtering out incomplete entries.
- **Encoding categorical variables**, such as user and movie IDs.
- **Splitting the dataset** into training (80%) and testing (20%) sets.
- **Normalizing ratings** to improve model performance.

Model Selection

Four different approaches were implemented:

1. Content-Based Filtering

This method recommends movies based on their **features**, such as genres. Similarity was measured using **Cosine Similarity** to find movies that are most similar to those a user has liked.

2. Collaborative Filtering

Collaborative filtering predicts user preferences based on past interactions.

- **User-User Collaborative Filtering:** Finds users with similar tastes and recommends movies liked by them.
- **Item-Item Collaborative Filtering:** Finds similar movies and recommends those based on a user's past interactions.
- Similarity measures used: **Cosine Similarity**

3. Matrix Factorization (SVD)

Singular Value Decomposition (SVD) was used to reduce the dimensionality of the user-movie interaction matrix. This method helps in extracting latent features from sparse data, improving recommendation quality.

4. Neural Collaborative Filtering (NCF)

Neural Collaborative Filtering was implemented using deep learning. The model consists of:

- **Embedding layers** to map users and movies to a latent space.
- **Fully connected layers (MLP)** to capture complex interactions.
- **Final output layer** to predict ratings.

The **NCF model was trained using the Adam optimizer** with **Mean Squared Error (MSE) loss**, ensuring efficient learning.

Key Insights

- **NCF outperformed other models** in terms of RMSE, precision, and recall.
- **SVD performed well** and is computationally efficient.
- **Item-Based Collaborative Filtering was slightly better** than User-Based Filtering.
- **Content-Based Filtering had limitations** as it does not consider user preferences dynamically.

3. Improvements & Future Optimizations

Current Improvements

- **Dropout layers were added** in NCF to prevent overfitting.
- **Hyperparameter tuning** improved model accuracy.
- **Matrix Factorization (SVD) provided a strong baseline** for evaluating deep learning models.

Future Enhancements

- **Hybrid Recommendation System:** Combining **content-based and collaborative filtering** for better results.
- **Autoencoders:** Using deep learning to uncover hidden patterns in user behavior.
- **Transformers and Attention Mechanisms:** Leveraging sequence models for improved recommendations.
- **Hyperparameter Optimization:** Using grid search for better learning rate and batch size tuning.