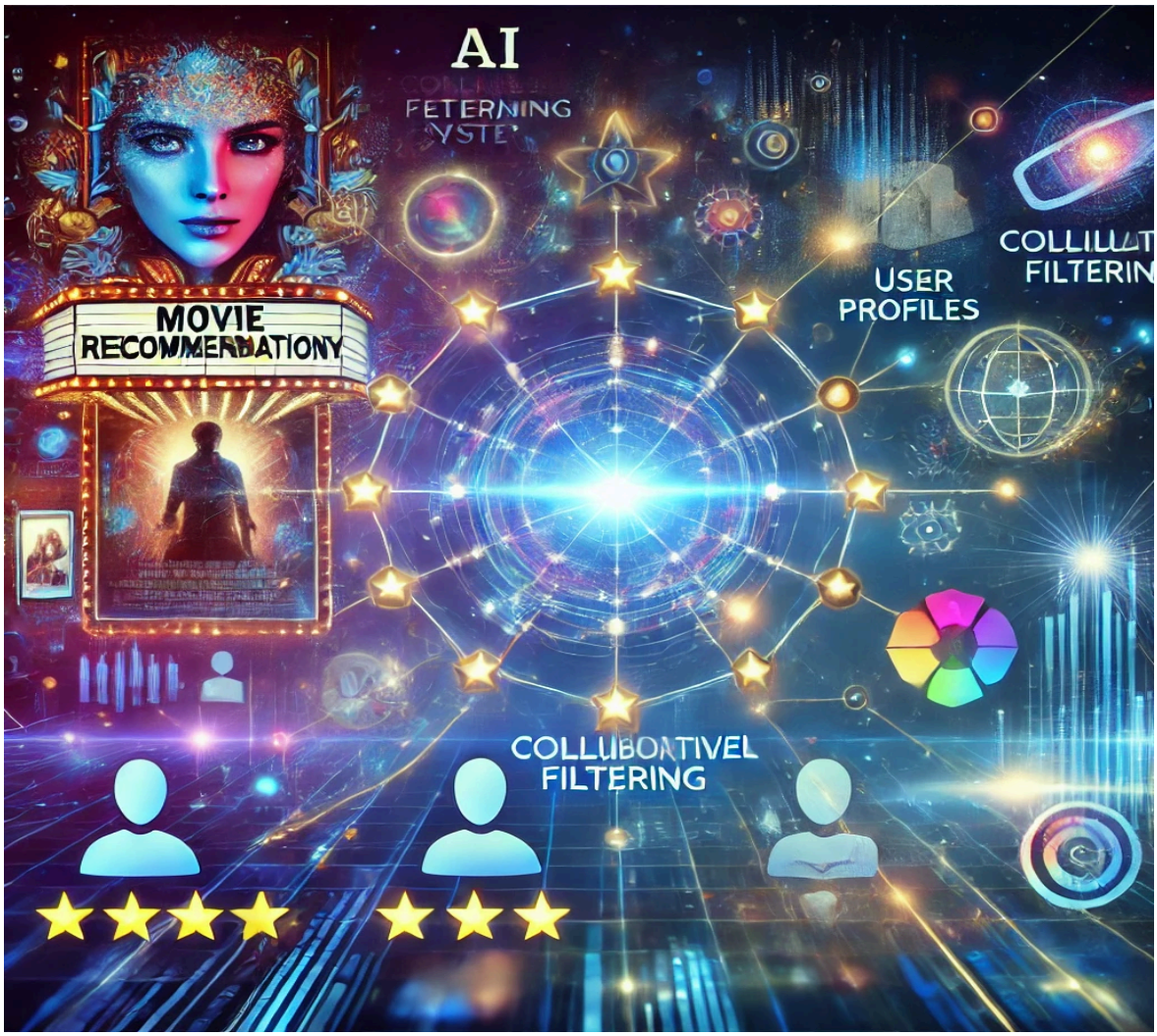
# Movie Recommendation System using Collaborative Filtering

**-Dinakaran.k**

**Overview:**




This project focuses on developing an AI-powered **Movie Recommendation System** using **Collaborative Filtering**. The system utilizes user rating data from the **MovieLens 20M Dataset** and applies **Alternating Least Squares (ALS)** to generate personalized movie recommendations.

## Key Features:

- **Collaborative Filtering Approach:** Uses user-item interactions to recommend movies.
- **MovieLens 20M Dataset:** Processes **20 million** movie ratings for accurate predictions.
- **Alternating Least Squares (ALS) Model:** Provides efficient matrix factorization for recommendations.
- **Sparse Matrix Representation:** Optimizes computation by storing only relevant ratings.
- **User-Specific Recommendations:** Suggests the top `N` movies based on past preferences.

## Workflow & Implementation:

#### 1. Data Preparation

- **Dataset Source:** Downloaded from **Kaggle (GroupLens MovieLens 20M dataset)**.
- **Files Used:**
    - `ratings.csv`: Contains `userId`, `movieId`, and `rating`.
    - `movies.csv`: Contains `movieId` and `title`.
- **Sparse Matrix Representation:**
    - Converts rating data into a **user-item matrix** to optimize storage and retrieval.

#### 2. Collaborative Filtering Using ALS

- **Algorithm:** `Alternating Least Squares (ALS)` for implicit feedback.
- **Training Process:**

- Factorizes the user-item matrix into **latent factors** for users and movies.
- Learns user preferences based on past ratings.
- Predicts scores for unseen movies.

**3. Movie Recommendation Generation**

- **Input:** User ID and the number of desired recommendations.
- **Process:**
  - Extracts top-rated unseen movies for the user.
  - Returns recommended movie titles along with predicted scores.

**4. Output & Interpretation**

**Example Input:**

```
Enter a User ID: 5
Enter number of recommendations: 5
```

**Example Output:**

```
Recommended Movies for User 5:
- Inception (Predicted Score: 4.75)
- The Dark Knight (Predicted Score: 4.72)
- Interstellar (Predicted Score: 4.69)
- Fight Club (Predicted Score: 4.67)
- The Matrix (Predicted Score: 4.65)
```

## Challenges and Solutions:

- **Challenge:** Large dataset size affecting model training speed.
    - **Solution:** Used **sparse matrix representation** to optimize memory usage.
- **Challenge:** Handling new users without historical ratings (cold start problem).
    - **Solution:** Future implementation can include **content-based filtering** for first-time users.
- **Challenge:** Improving recommendation diversity.
    - **Solution:** Introduced **regularization** in ALS to reduce overfitting and recommend diverse content.

## Progress and Next Steps:

**Accomplishments:**

- Successfully implemented **Collaborative Filtering** with **ALS**.
- Processed **20M+ ratings** from the MovieLens dataset efficiently.
- Developed a **user-specific recommendation system** with real-time inputs.

## Next Steps:

- Integrate **Hybrid Recommendation System** (combine **Collaborative & Content-Based Filtering**).
- Implement a **web interface** for user-friendly recommendations.
- Improve model scalability by using **GPU-accelerated ALS training**.

## Conclusion:

The **Movie Recommendation System using Collaborative Filtering** efficiently predicts user preferences based on past ratings. The **ALS model** provides accurate and scalable recommendations, making it ideal for streaming platforms, entertainment services, and personalized content discovery.