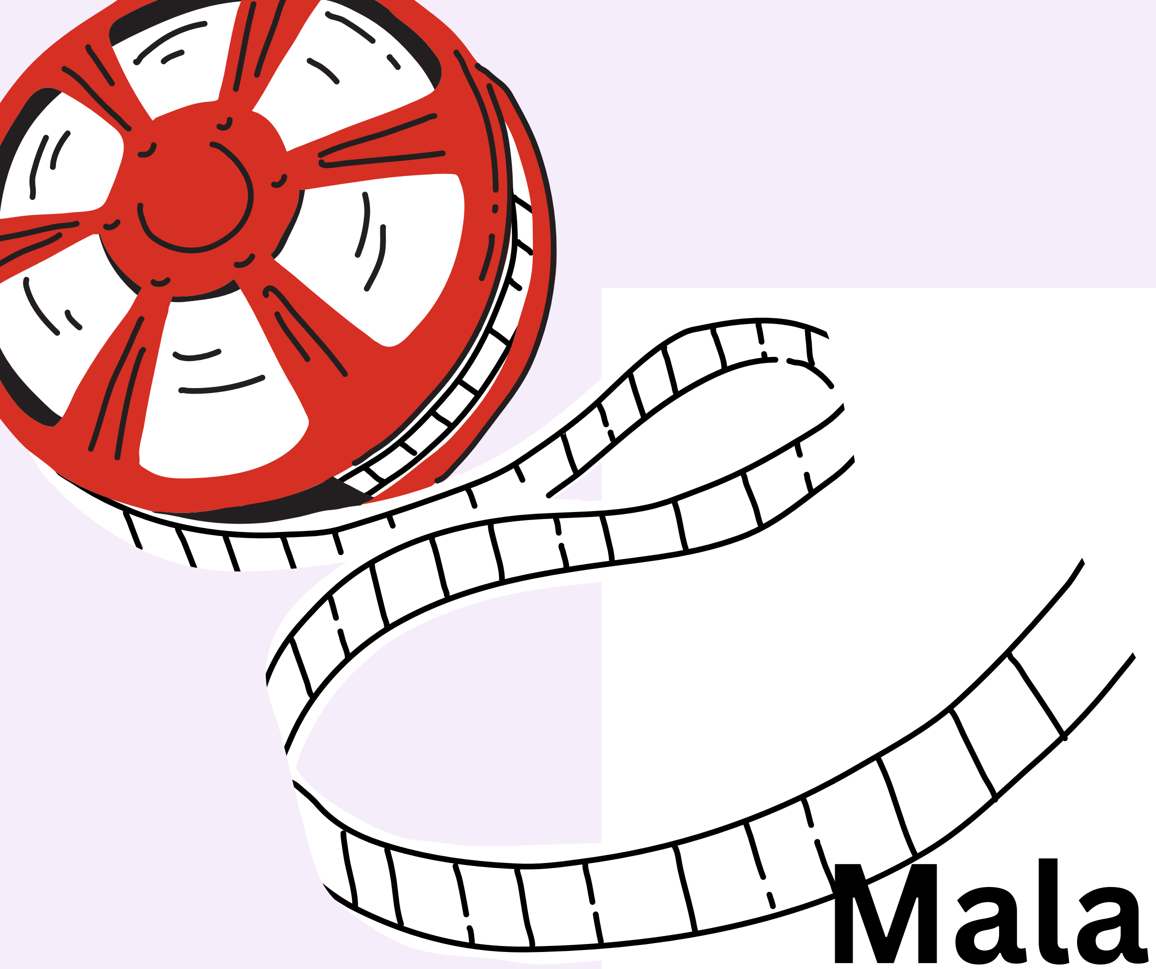
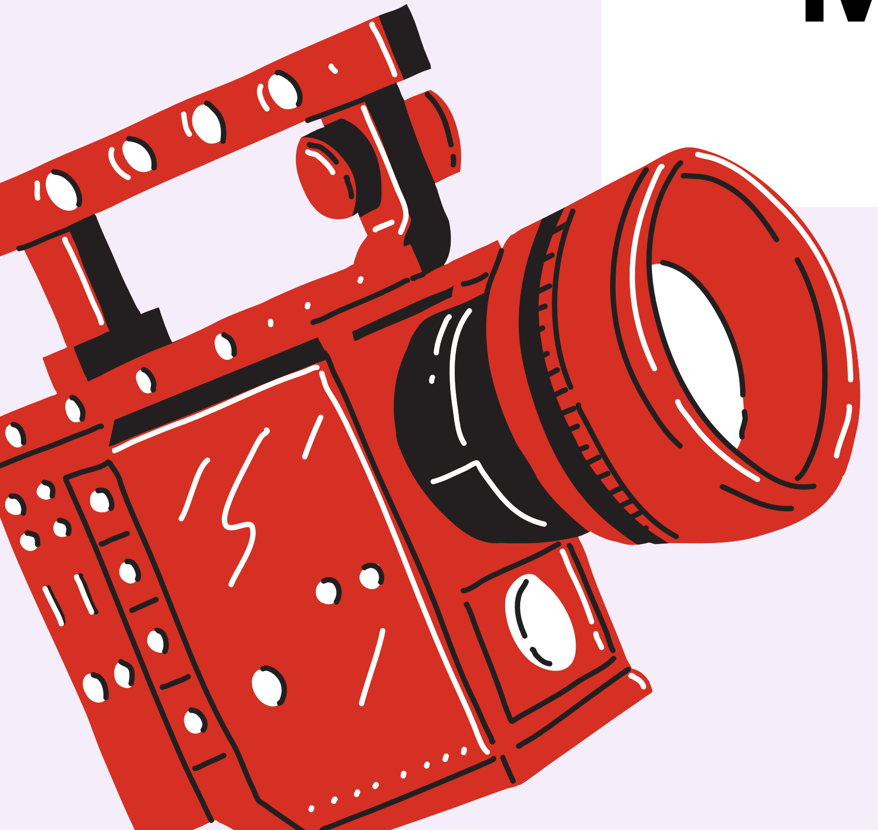


Movie Recommendation System Project

Malak Mohamed Ragab 2305303
Logina Mahmoud Elhag 2305532
Menna Walid Mahmoud 2305303



Introduction

1

In the era of digital content consumption, recommendation systems play a vital role in personalizing user experiences. This project aims to build a Movie Recommendation System using three approaches:

- Content-Based Filtering
- Collaborative Filtering
- Hybrid Recommendation Model

These systems suggest movies to users based on their preferences and past interactions, enhancing engagement and satisfaction.

Datsaset description

We used two datasets:

✅ movies.csv

Contains movie metadata such as:

- Movie ID
- Title
- Genres (e.g., Action, Comedy)

✅ ratings.csv

Contains user-generated ratings with:

- User ID
- Movie ID
- Rating (0.5 to 5)
- Timestamp

These datasets were preprocessed and merged for effective modelin

Data preprocessing

Main steps include:

- Cleaning missing/null values
- Converting timestamps (optional for time-aware models)
- Encoding genres using TF-IDF or one-hot
- Creating a user-item interaction matrix

📌 Efficient preprocessing ensures quality inputs for better prediction

Why learn about them?

Understanding the language of film posters can help us develop critical thinking skills, enhance our appreciation of film as an art form and help us become more discerning viewers.

While the elements of film posters can vary depending on the style and genre of the film there are a collection of features commonly found in most film posters. Let's take a look at them...

Content-Based Filtering

Recommends movies similar to those the user liked, based on movie features.

Techniques used:

- TF-IDF on genres and descriptions
- Cosine similarity between movie vectors
- Personalized recommendations without relying on other users

🔄 Works well for new or inactive users, but may suffer from limited discovery

Collaborative Filtering

Predicts user preferences by learning from the ratings of similar users.

Techniques used:

- User-User similarity (based on rating patterns)
- Item-Item similarity
- SVD (Singular Value Decomposition) from Surprise library

🤝 Excels when there's sufficient user rating data. Suffers from the "cold start" problem.

Hybrid Recommendation

Combines strengths of both models to overcome their weaknesses.

Advantages:

- Uses content-based for new users/items
- Uses collaborative filtering when user history is sufficient
- Combines both using a weighted strategy

🧠 Offers higher flexibility, accuracy, and user satisfaction.

An intuitive web interface using Streamlit.

Features:

- Enter a movie name or user ID
- View top 10 movie recommendations
- Clean, minimal design for better user experience
- Option to switch between recommendation types (if extended)

💡 Easy to demo, deploy, and interact with.

Libraries & Tools Used

Purpose	Library
Data manipulation	`pandas`
Numerical operations	`numpy`
Vectorization, similarity, ML tools	`scikit-learn`
Collaborative filtering & SVD	`surprise`
Web app interface	`streamlit`
Optional: Data visualization	`matplotlib / seaborn`

```python
import pandas as pd
```

python

```python
import numpy as np
```

python

```python
from sklearn.feature_extraction.text import TfidfVecto
from sklearn.metrics.pairwise import cosine_similarity
```

python

```python
from surprise import SVD, Dataset, Reader
from surprise.model_selection import train_test_split
```

python

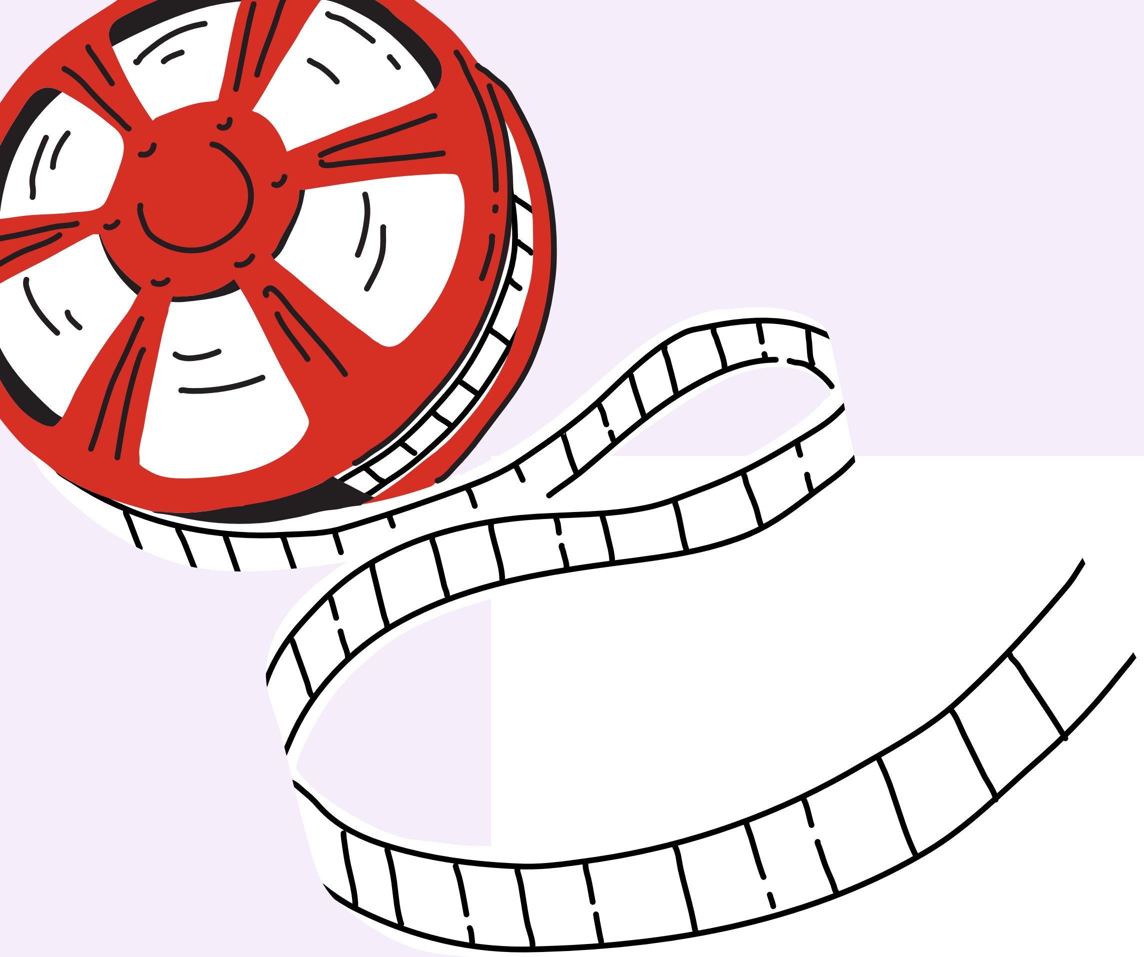
```python
import streamlit as st
```

Challenges & Solutions

Solution	Challenge
Used matrix factorization (SVD)	Sparse ratings matrix
Used content-based filtering	Cold start for new users
Applied preprocessing techniques	Data cleaning
Optimized with sampling and modular code	Performance on large datasets

Conclusion

This project illustrates how to build a full recommendation engine from scratch using real-world data. By combining content-based and collaborative filtering into a hybrid system, we achieved better precision, diversity, and scalability.
The project reflects real challenges in building recommender systems and serves as a foundation for more advanced, real-time, and large-scale solutions in the future



the end