Movie Recommendation Engine - Beginner Guide

Introduction

We're building a simple system that recommends movies similar to a movie the user likes.

Like how Netflix shows 'Because you watched X...'

Libraries Used

Libraries Used:

- pandas: To read and work with CSV files (data handling)
- sklearn.metrics.pairwise.cosine_similarity: To find similarity between movies

Step-by-Step Code Explanation

Step 1: Load the CSV Files

```
ratings = pd.read_csv("ratings.csv")
movies = pd.read_csv("movies.csv")
```

Step 2: Merge the Two Files

```
data = pd.merge(ratings, movies, on='movieId')
```

Step 3: Create a User-Movie Rating Table

```
user_movie_matrix = data.pivot_table(index='userId', columns='title', values='rating')
```

Step 4: Fill Missing Ratings with 0

```
user_movie_matrix.fillna(0, inplace=True)
```

Step 5: Calculate Similarity Between Movies

```
from sklearn.metrics.pairwise import cosine_similarity
similarity_matrix = cosine_similarity(user_movie_matrix.T)
```

Step 6: Convert Similarity Matrix to DataFrame

```
movie_similarity_df = pd.DataFrame(similarity_matrix, index=user_movie_matrix.columns, columns=user_movie_matrix.columns)
```

Step 7: Create the Recommendation Function

```
def recommend_movies(movie_title, top_n=5):
    if movie_title not in movie_similarity_df.columns:
        return f" Movie '{movie_title}' not found in dataset."
    similarity_scores = movie_similarity_df[movie_title].sort_values(ascending=False)
    return similarity_scores[1:top_n+1].index.tolist()
```

Step 8: Run It!

```
movie_to_search = "Toy Story (1995)"
recommendations = recommend_movies(movie_to_search)
print(f"Recommendations for '{movie_to_search}':")
for i, rec in enumerate(recommendations, 1):
    print(f"{i}. {rec}")
```

Sample Output

Sample Output:

```
Recommendations for 'Toy Story (1995)':
```

1. Bug's Life, A (1998)
2. Aladdin (1992)
3. Beauty and the Beast (1991)
4. Lion King, The (1994)
5. Hercules (1997)

Conclusion

Congratulations! You've built a simple content-based movie recommendation engine using user rating data.

You can improve this project by:

- Adding genre-based filtering
- Using collaborative filtering with surprise or LightFM
- Building a web app interface using Streamlit

Best of luck with your project!