Movie Analytics & Recommendation System

A Netflix-Style Analytics & Recommendation Prototype

This project is a MovieLens-based analytics and recommender system developed as part of a final assessment for a scenario involving a streaming technology company. The system delivers business insights and deployable recommendation features using machine learning.

Project Objectives

- Analyse large-scale movie ratings and metadata
- Extract actionable business insights
- Build and evaluate multiple recommendation models
- Deploy an interactive analytics and recommendation application
- Demonstrate business value for content strategy and user retention

Recommendation Models Implemented

1. **Model**
    - Weighted popularity baseline, rigde, logistic regression, random forest, Knn, gradient boost, decision tree)
    - Suitable for trending and homepage recommendations
2. **Content-Based Filtering**
    - Genre and tag similarity
    - Enhances content discovery
3. **Collaborative Filtering**
    - User–user similarity
    - Personalised recommendations

Application Features

- Interactive analytics dashboard
- Movie similarity recommendations
- User-based recommendations
- Fixed-size movie cards for consistent UI
- Graceful handling of missing posters

Tech Stack

- **Language:** Python
- **Framework:** Streamlit
- **Libraries:**
  - ✓ Pandas
  - ✓ NumPy
  - ✓ Scikit-learn
  - ✓ Plotly

Running the Application

1. Run the App

streamlit run my_app.py

📁 **Project Structure**

```
├── ml-latest
├── ml-latest/
│   ├── movies.csv
│   ├── ratings.csv
│   ├── links.csv
│   ├── genome_tags.csv
│   ├── genome_scores.csv
│   └── tags.csv
├── output/
│   ├── plots.csv
│   ├── reports.csv
│   ├── models.csv
│   │   ├── baseline_model.pkl, lr.pkl, lasso.pkl, gb.pkl, rf.pkl, dt.pkl, knn.pkl
├── dashboard/
│   ├── MovieLens_Dashboard_Data/
└── README.md
```

Business Value

- Supports content acquisition decisions
- Improves user engagement and retention
- Demonstrates deployable ML for real-world streaming platforms

Ethical Considerations

- Anonymised user data
- Bias-aware popularity scoring
- Multiple recommendation strategies for fairness

Future Work

- Hybrid recommendation engine
- Real-time user interaction tracking
- Cloud deployment
- A/B testing of recommendation strategies