ivy-chelagat / Recommender-System---Movies

Code Issues Pull requests Actions Projects Wiki Security Insights Settin

0 stars 0 forks 0 watching Branches Activity
Tags

Public repository

4 Branches 0 Tags Go to file t Go to file Add file Code

ivy-chelagat Fix punctuation and spacing in README.md 5d8262b · 4 minutes ago

| | | |
|---|---|---|
| Recommender_System_files | Final README with working image ... | 2 days ago |
| .gitignore | Create .gitignore | 4 days ago |
| Data Report.pdf | Add notebook for genre average ra... | 2 days ago |
| README.md | Fix punctuation and spacing in REA... | 4 minutes ago |
| Recommender System .ipynb | Add notebook for genre average ra... | 2 days ago |
| links.csv | font changes and cell corrections | 5 days ago |
| ml-latest-small.zip | Add files via upload | 5 days ago |
| movies.csv | font changes and cell corrections | 5 days ago |
| notebook.pdf | initial commit | 2 hours ago |
| presentation.pdf | Rename Presentation.pdf to presen... | 3 hours ago |
| ratings.csv | font changes and cell corrections | 5 days ago |

README

# Business Overview

This project focuses on building a movie recommendation system inspired by platforms like Netflix and Showmax. It aims to tackle the challenge of “scroll fatigue” , when users spend too much time deciding what to watch by providing personalized movie recommendations based on user preferences and behavior.

Using the MovieLens dataset, the project develops and evaluates both Collaborative Filtering (SVD) and Hybrid (SVD + Content-Based) models to predict user ratings and generate a top-five list of personalized movie suggestions.

# Stakeholders

**Streaming Platforms (e.g., Netflix, Showmax):** Want to increase user engagement and retention.

**Users:** Seek quick, accurate, and relevant movie recommendations.

**Data Science Teams:** aim to design efficient, accurate, and scalable recommendation systems.

# Key Business Questions

How can user viewing patterns and preferences be used to recommend movies they’re likely to enjoy?

Which recommendation technique, collaborative filtering or hybrid, offers the best balance between accuracy and diversity?

How much can a personalized recommendation system improve user satisfaction and retention for streaming services?

# Business Objectives

Develop a recommendation model that predicts user ratings for unseen movies.

Evaluate models using RMSE and MAE metrics.

Generate top-5 movie recommendations for each user.

Demonstrate how the system enhances personalization and engagement.

# Success Criteria

Achieve RMSE ≤ 0.90 and MAE ≤ 0.70 on test data.

Generate meaningful, personalized recommendations for users.

# Data Understanding

**Source of Data**

The project uses the MovieLens Small Dataset, which contains:

100,000 ratings

600+ users

9,000+ unique movies Each record includes a user ID, movie ID, rating, timestamp, and movie metadata such as title and genres

**Libraries and Tools Used**

Pandas, NumPy, Seaborn, Matplotlib – Data analysis & visualization

Surprise – Collaborative Filtering (SVD) implementation

Scikit-learn – TF-IDF, Cosine Similarity, evaluation metrics

Pickle – Model serialization

# Model Overview

**Collaborative Filtering (SVD)**

- Built using the Surprise library's SVD algorithm.
- Hyperparameter tuning via GridSearchCV to optimize learning rate, regularization, and latent factors.
- Achieved **RMSE = 0.87** and **MAE = 0.67** — meeting success criteria.

**Content-Based Filtering (CBF)**

- Constructed using TF-IDF and cosine similarity on movie genres.
- Recommends movies similar in content to those a user already liked.

**Hybrid Model**

- Combines SVD and CBF:

  Final Score = $\alpha$ × SVD Score + (1 − $\alpha$) × Genre Score

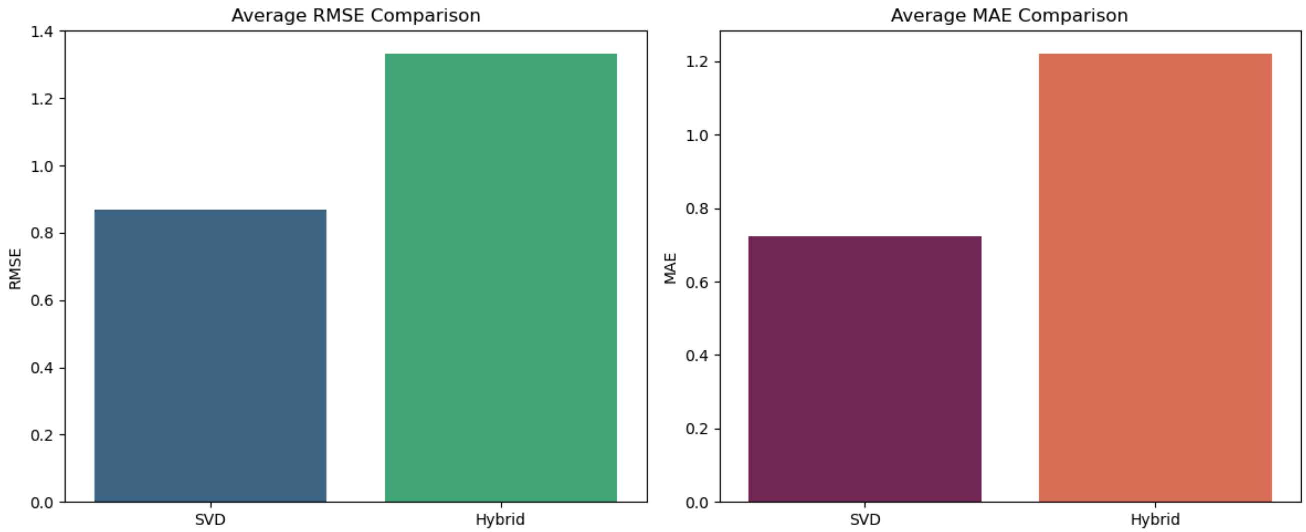
- Balances personalization (SVD) with diversity (genre similarity).
- Improves the ranking quality of top recommendations despite slightly higher RMSE.

# Model Evaluation

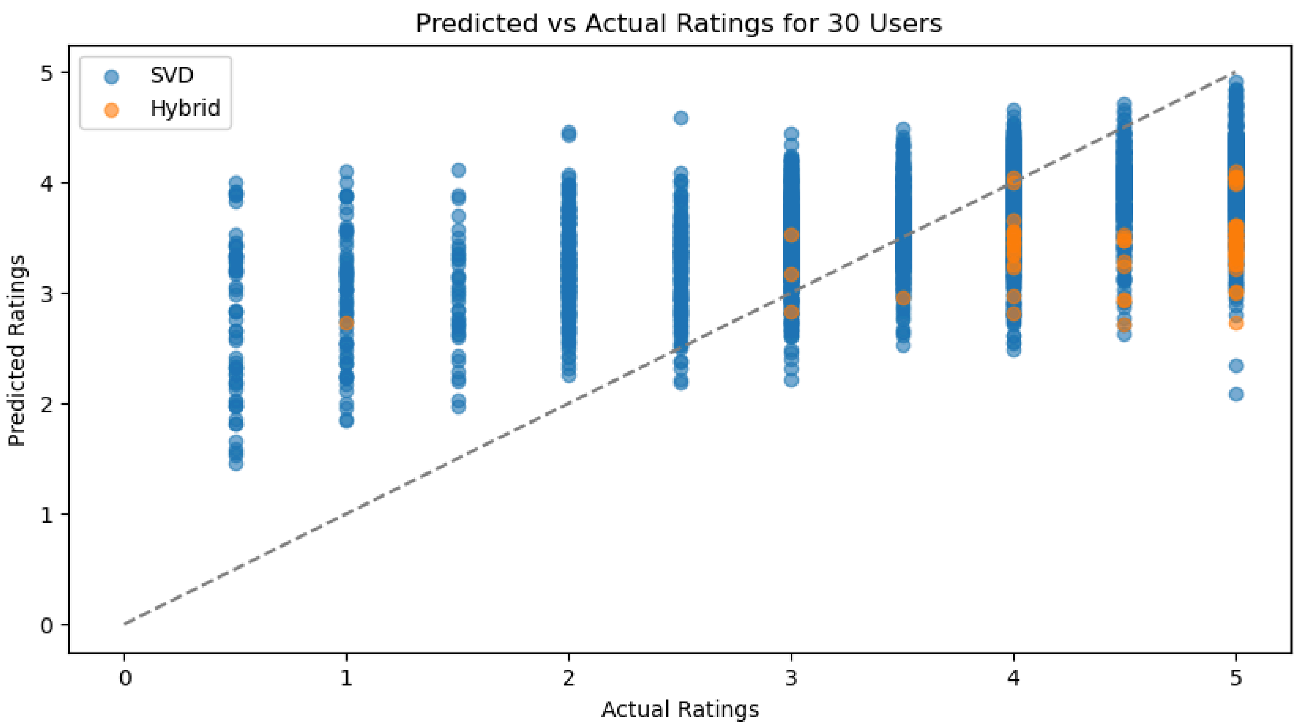
| Model | RMSE | MAE | Notes |
|---|---|---|---|
| **SVD (Collaborative)** | 0.86 | 0.67 | Accurate, strong baseline |
| **Hybrid (SVD + Genre)** | 1.30 | 1.20 | Slightly higher error but more diverse recommendations |

# Key Evaluation Visuals

Bar Chart: Compared average RMSE and MAE between SVD and Hybrid models.




Scatter Plot: Showed predicted vs. actual ratings — Hybrid predictions clustered more tightly for highly rated movies.




# Conclusion

The system successfully predicts user preferences and recommends relevant movies.

The SVD model achieved strong accuracy, while the Hybrid model improved diversity and personalization.

Personalized recommendations enhance user satisfaction and can reduce “scroll fatigue” in streaming platforms.

# Key Findings

Collaborative Filtering (SVD) effectively learns latent features, achieving an average RMSE < 0.9.

Hybridization improves the quality of top-N recommendations by blending accuracy and novelty.

Most users favor Drama, Animation, and Documentary genres — useful for content acquisition strategies.

# Recommendations

1. Deploy the Hybrid Model in production to balance personalization and discovery.
2. Enrich data with additional metadata (tags, actors, demographics) to improve precision.
3. Continuously retrain the model with new user ratings to maintain freshness.
4. Implement user feedback loops for adaptive learning.
5. Integrate into a web or mobile interface for real-time interaction.

# Next Steps

Develop a weighted user profile capturing individual genre preferences.

Experiment with Deep Matrix Factorization or Neural Collaborative Filtering (NCF).

Evaluate coverage and diversity metrics to ensure recommendations remain varied.

### Releases

No releases published
Create a new release

### Packages

No packages published
Publish your first package

### Contributors 5

### Languages

- Jupyter Notebook 100.0%