**Movie recommendation system - final report**

**Authors: Julia Biwan, Weronika Wódkowska**

**link to huggingface:** [**https://huggingface.co/JuliaBiwan/movie\_recomendations**](https://huggingface.co/JuliaBiwan/movie_recomendations)

**Objectives**

| **Goal** | **Description** |
| --- | --- |
| Predict ratings | To build a model that estimates movie ratings based on other users’ preferences. |
| Personalized suggestions | To create a system that recommends movies users might enjoy. |
| Checking the accuracy | To measure accuracy using RMSE (Root Mean Squared Error). |

**Milestone 1: Data collection and preprocessing**

| **Source** | **IMDb dataset downloaded from Kaggle** |
| --- | --- |
| number of movies | 1,000 |
| chosen features | title, genre, director, rating |

Cleaning and preprocessing

| **Step** | **What did we opt for** |
| --- | --- |
| Selected key columns | Used only relevant info such as title, genre, director and rating. |
| Handled missing data | Filled missing genres/directors with "Unknown". |
| Removed duplicates | Dropped duplicate movies to ensure clean data. |
| Simulated ratings | Created fake ratings for 100 movies across 3 users. |

**Milestone 2: Model development and training**

Models used

| **Model** | **Method** |
| --- | --- |
| SVD | Breaks down ratings into underlying factors. |
| User-Based CF | Finds similar users to recommend movies. |
| Item-Based CF | Recommends movies based on similar items |

Training and testing data

| **Dataset Split** | **Percentage** |
| --- | --- |
| Training Set | 80% |
| Testing Set | 20% |

Model performance (RMSE scores)

| **Model** | **RMSE** |
| --- | --- |
| SVD | 1.4254 |
| User-Based CF | 1.9012 |
| Item-Based CF | 1.4228 |

Best Model: Item-based collaborative filtering with 1.4228 RMSE.

**4. Milestone 3: Model evaluation and final report**

Final model comparison

| **Model** | **Accuracy** |
| --- | --- |
| SVD | Decent predictions |
| User-Based CF | Least accurate |
| Item-Based CF | Best accuracy |

How it works

1. Users enter movies they like.
2. The system finds similar movies using:

* Collaborative Filtering (User-Based, Item-Based, SVD).
* Content Filtering (Genres, Directors).

Example test run

| **User input** | **Recommended movies** |
| --- | --- |
| 'Magnolia', 'Dogville', 'The Incredibles' | Boogie Nights  There Will Be Blood  Mononoke-hime  Dancer in the Dark  The Machinist |

Final answer - How effective is the model?

| **Metric** | **Value** |
| --- | --- |
| Best RMSE Score | 1.4228 |
| Most effective model | Item-based CF |
| Accuracy | High - provides relevant recommendations |

→ Our model makes personalized recommendations with strong accuracy.

5. Next steps and improvements

| **Improvement** | **Description** |
| --- | --- |
| More real user data | Replace fake ratings with actual user feedback |
| Deep learning | Explore neural networks for better predictions |
| Better content filtering | Improve genre-based suggestions |
| Model upload to Hugging Face | Publish model with dataset details, RMSE score, and training info. |

**Final comment**

In this project, we built a movie recommendation system using collaborative filtering and content-based filtering. The best-performing model (Item-Based CF) achieved an RMSE of 1.4228, showing strong accuracy in predicting user preferences.

Final output: A system that suggests personalised movie recommendations based on user input and similarity scores.