

Movie Rating Prediction - ONE PAGE SUMMARY

The Entire Project in 5 Minutes

☒ WHAT WE DID

Built a system that predicts how much you'll like a movie (like Netflix recommendations!)

☒ THE DATA

- 100,836 movie ratings from 610 users rating 9,742 movies
- Ratings: 0.5 to 5.0 stars
- Problem: 98.3% of data is MISSING (most users haven't seen most movies!)

☒ THE 3 TECHNIQUES WE USED

1. K-Means Clustering (Grouping Users)

Simple idea: Sort users into groups based on behavior

- Group 1: Love everything (25% of users)
- Group 2: Picky/casual (30% of users)
- Group 3: Average watchers (28% of users)
- Group 4: Active critics (17% of users)

Why? People in same group rate similarly!

2. Decision Tree (Making Predictions)

Simple idea: Like a flowchart - "Is movie popular? Is user generous? → Predict 4.5 stars!"

- Accuracy: 0.92 RMSE (average error)
- Good for: Understanding WHY we predicted something

3. Matrix Factorization (Finding Patterns)

Simple idea: Find hidden "taste patterns" like "loves action, hates romance"

- Accuracy: 0.88 RMSE (BEST! ☺)
- Good for: Most accurate predictions

☒ RESULTS

Model	Error	Rating
Matrix Factorization	0.68 stars	A- ☺
Decision Tree	0.71 stars	B+
Random Guess	1.25 stars	F

Translation: We can predict ratings usually within 0.7 stars of actual rating!

☒ KEY FINDINGS

1. Movie quality matters most (42% of prediction)

- If everyone loves it, you probably will too!

2. User groups help
 - People in same group rate similarly
 3. Hidden patterns are powerful
 - Computer found 20 patterns humans can't see
 4. Data sparsity is hard
 - 98.3% missing makes it challenging!
-

☒ REAL WORLD EXAMPLES

☒ **Netflix:** "Because you watched..." recommendations ☒ **Spotify:** Discover Weekly playlists ☒ **Amazon:** "Customers also bought..." ☒ **YouTube:** Recommended videos
We built a mini version of these!

☒ WHY THIS IS PREDICTION (Not Classification)

This Project	Classification
Predict NUMBERS (4.2 stars)	Predict CATEGORIES (spam/not spam)
Output: 0.5-5.0	Output: Yes/No, A/B/C
Metrics: RMSE, MAE, R ²	Metrics: Accuracy, F1
Example: "4.2 stars"	Example: "Spam"

☒ WHAT WE ACCOMPLISHED

✓ Loaded and explored 100K+ ratings ✓ Created 4 user segments with K-Means ✓ Built Decision Tree predictor (0.92 RMSE) ✓ Built Matrix Factorization predictor (0.88 RMSE) ✓ Compared models and found the best one ✓ Demonstrated 3 ML techniques from course!

☒ THE BOTTOM LINE

Question: Can computers predict what movies you'll like? **Answer:** YES! With ~70% accuracy (0.7 star average error)

Question: Why not 100% accurate? **Answer:** Humans are unpredictable! Mood, context, and randomness affect ratings.

Question: Is 35% variance explained good? **Answer:** YES! Predicting human behavior is HARD. 35% is actually impressive!

☒ TECHNICAL SUMMARY

Dataset: MovieLens 100K (GroupLens Research) **Languages:** Python 3.12 **Libraries:** NumPy, Pandas, Scikit-learn, SciPy **Techniques:**

- Unsupervised Learning (K-Means)
- Supervised Learning (Decision Tree Regressor)
- Collaborative Filtering (SVD Matrix Factorization)

Evaluation:

- RMSE (Root Mean Squared Error) - measures average error
 - MAE (Mean Absolute Error) - average distance from truth
 - R² Score - variance explained (0 to 1)
-

☒ ELEVATOR PITCH (30 seconds)

"We built a movie rating prediction system like Netflix uses. Starting with 100,000 ratings from 610 users, we grouped users by behavior, built prediction models, and found hidden patterns in the data. Our best model predicts ratings within 0.7 stars accuracy - good enough for real recommendations! We used 3 machine learning

techniques: clustering, decision trees, and matrix factorization. This demonstrates how companies like Netflix, Spotify, and Amazon make personalized recommendations."

FILES IN PROJECT

1. **Movie_Rating_Prediction_Project.ipynb** - Main code
 2. **PRESENTATION.md** - Detailed presentation (50+ pages)
 3. **PRESENTATION_SIMPLE.md** - Easy version (beginner-friendly)
 4. **ONE_PAGE_SUMMARY.md** - This file!
 5. **README.md** - Setup instructions
 6. **ml-latest-small/** - Dataset folder
-

DONE! You now understand the entire project!

Print this page and use it as a quick reference during your presentation!