Movie Recommendation System

By: Mohamed Kallel
Simon-Pierre Rodner
Jean-Christophe Rigoni

movielens

Introduction

- OBJECTIVE : to recommend movies
- DATA EXPLORATION
- MODEL BUILDING AND EVALUATION
- CONCLUSION

Objectives: Recommend movies the right way

It needs to be segmented into different use cases

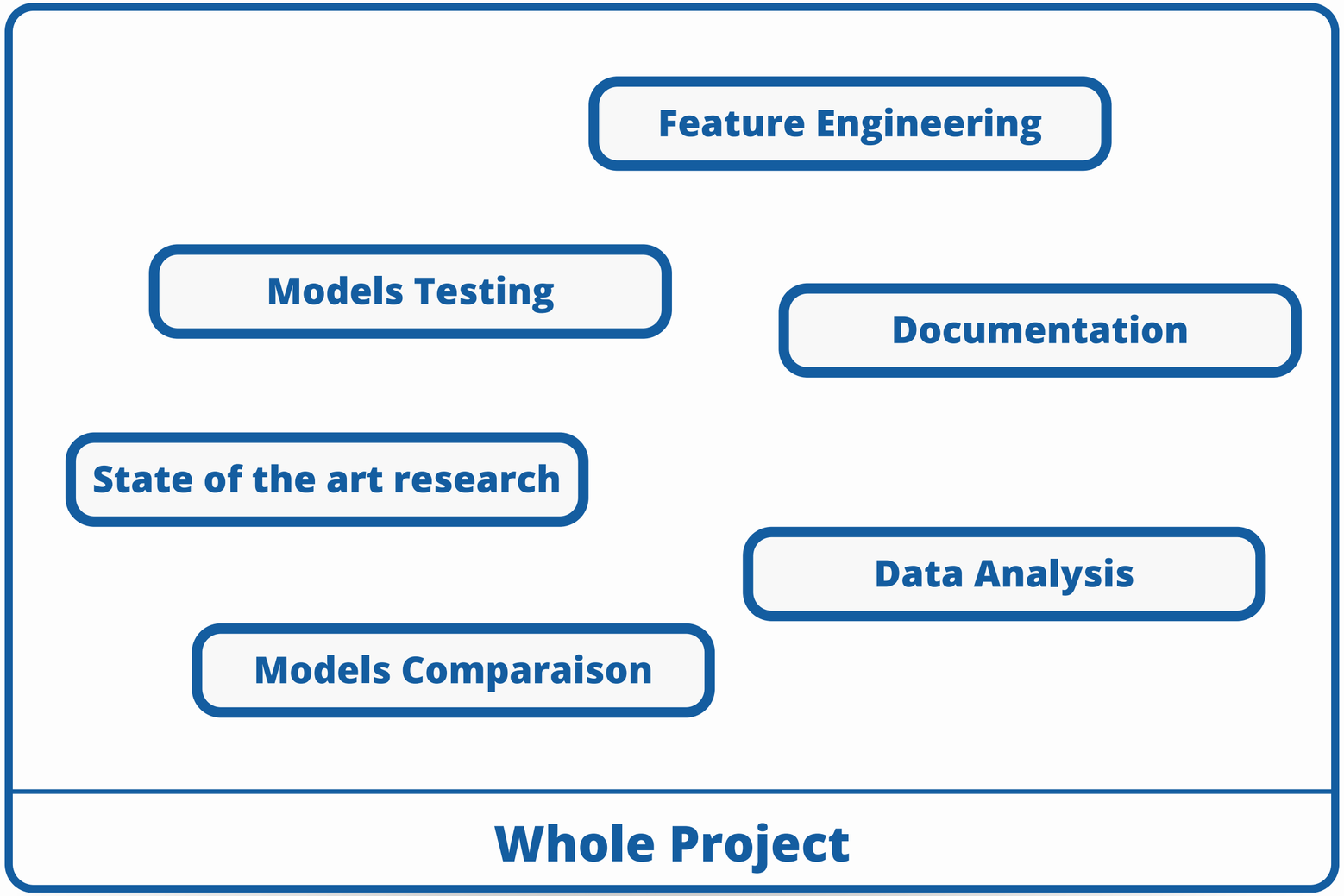
01 Estabished Users and Movies
Collaborative Filtering

02 Because you watch this movie
Hybrid Collaborative-Content

03 ML with Filtered data:
Genre - Popularity - Year

04 Cold Start : Content-Based
Rating and Popularity

Project Planning



Project Planning

State of the art research

Data Analysis

Feature Engineering

Models Testing

Models Comparaison

Documentation

Whole Project

Project Planning

- State of the art research
- Data Analysis
- Feature Engineering
- Models Testing
- Models Comparaison
- Documentation

Whole Project

Project Planning

Whole Project

- Feature Engineering
- Models Testing
- Models Comparaison
- Documentation

Weekly Planning

- State of the art research
- Data Analysis

Project Planning

Whole Project

- Feature Engineering
- Models Testing
- Models Comparaison
- Documentation

Weekly Planning

- State of the art research
- Data Analysis

Weekly Meeting

- Progress Status
- Learnt Information
- Problems/Bottlenecks
- Next Steps

DATASET EXPLORATION

- **rating**: usersID - movieId - ratings - date
- **movie** : movieId - title (year) - genres
- **tag** : usersID - movieId - tag - date
- **genome_scores** : movieId - tagId - relevance
- **genome_tags** : tagId - tag
- **link**: movieId - IMDb - TMDb

Ratings Distribution

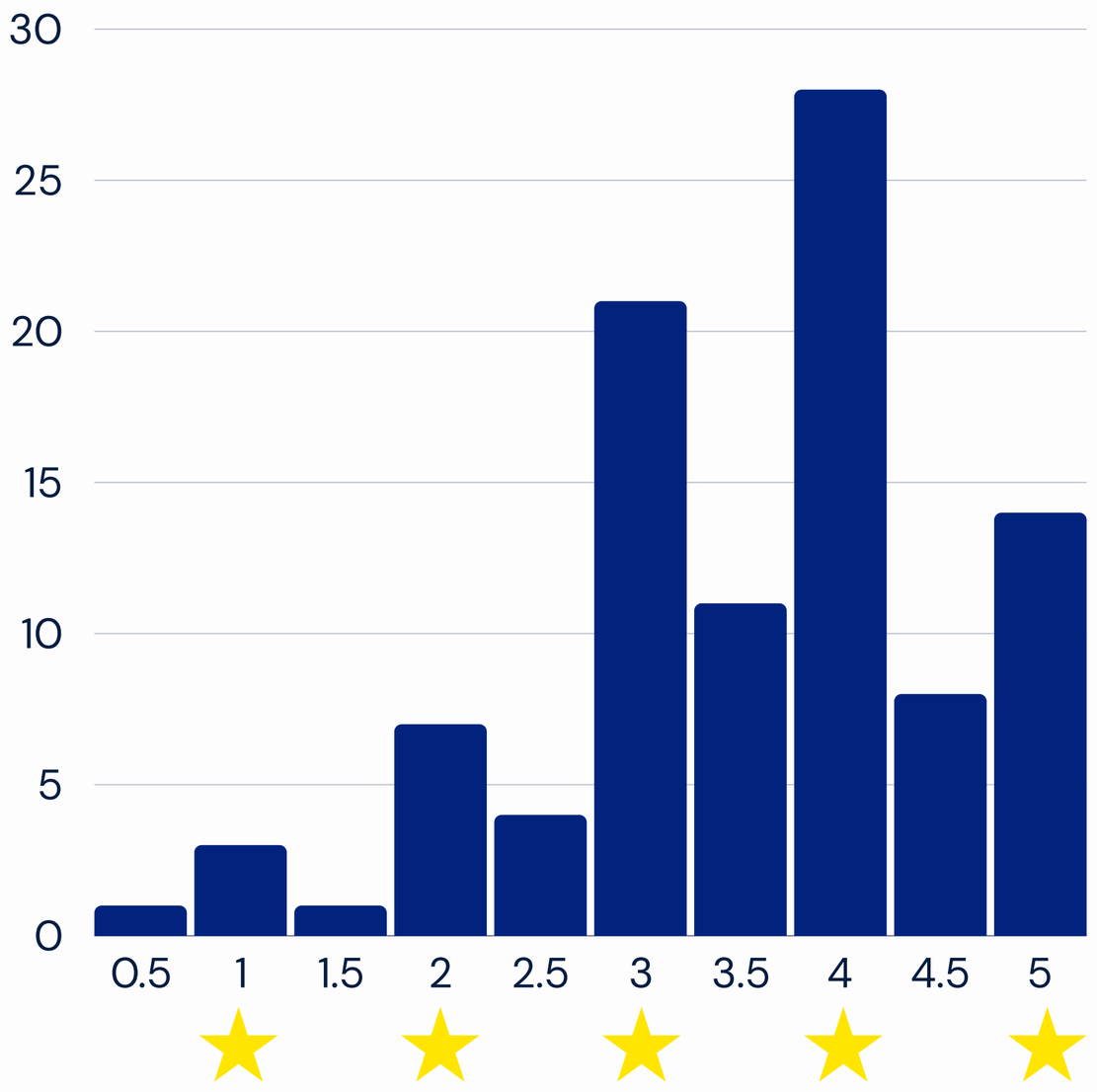
Most users give 3 stars and above.

3.53
AVERAGE RATING

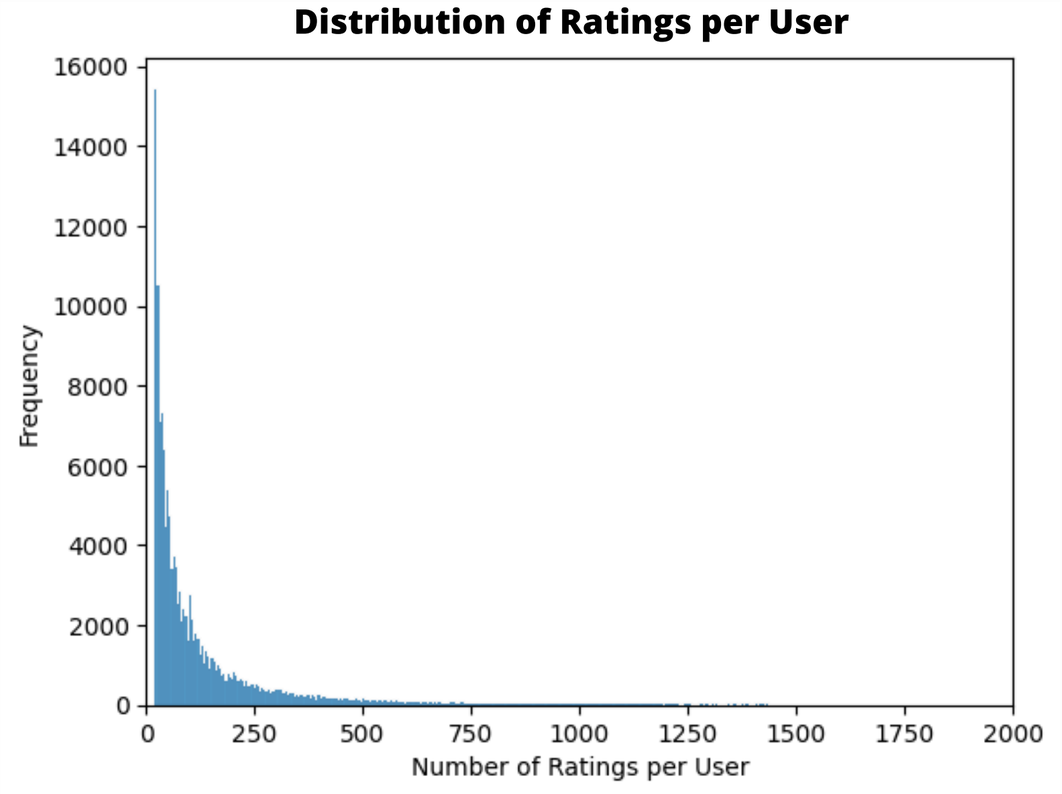
20M
NUMBER OF RATINGS

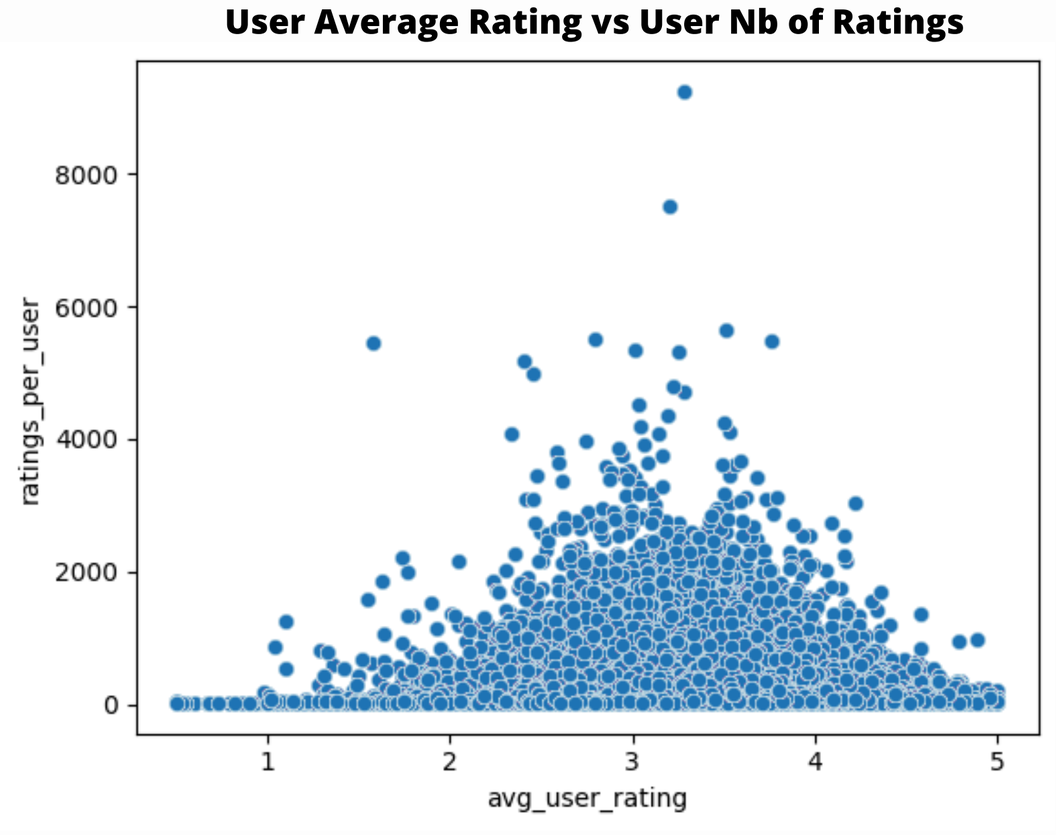
138K
NUMBER OF USERS

27K
NUMBER OF MOVIES

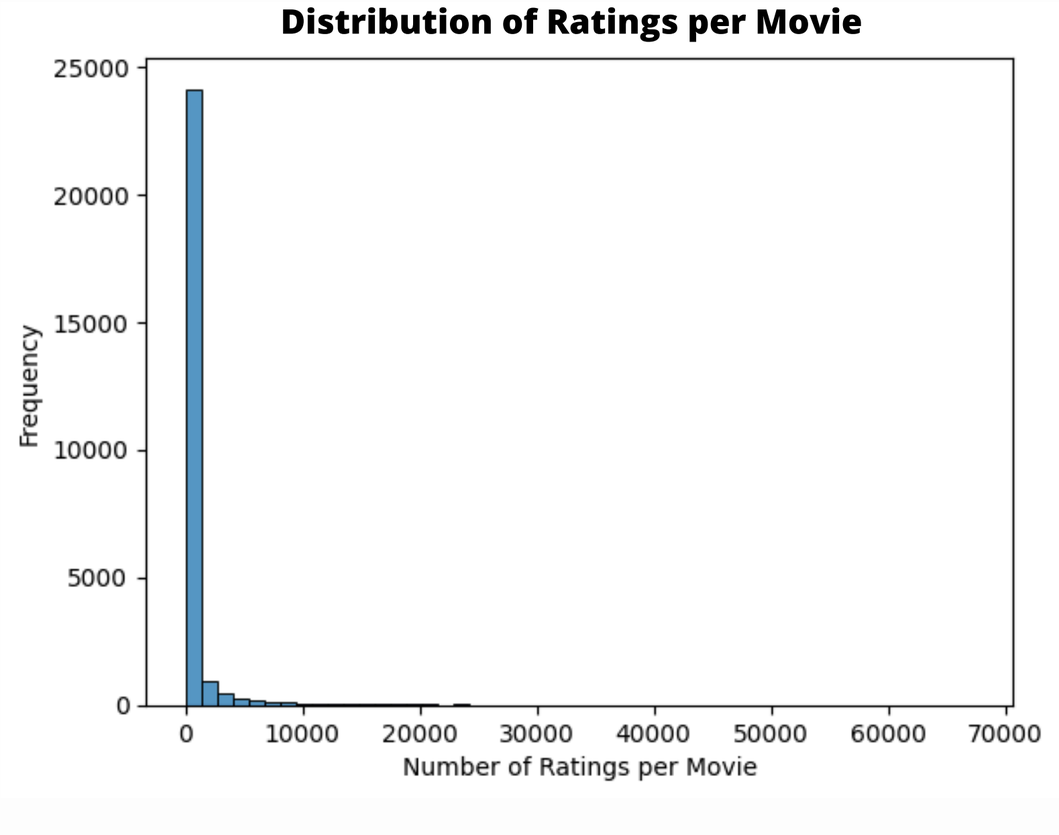


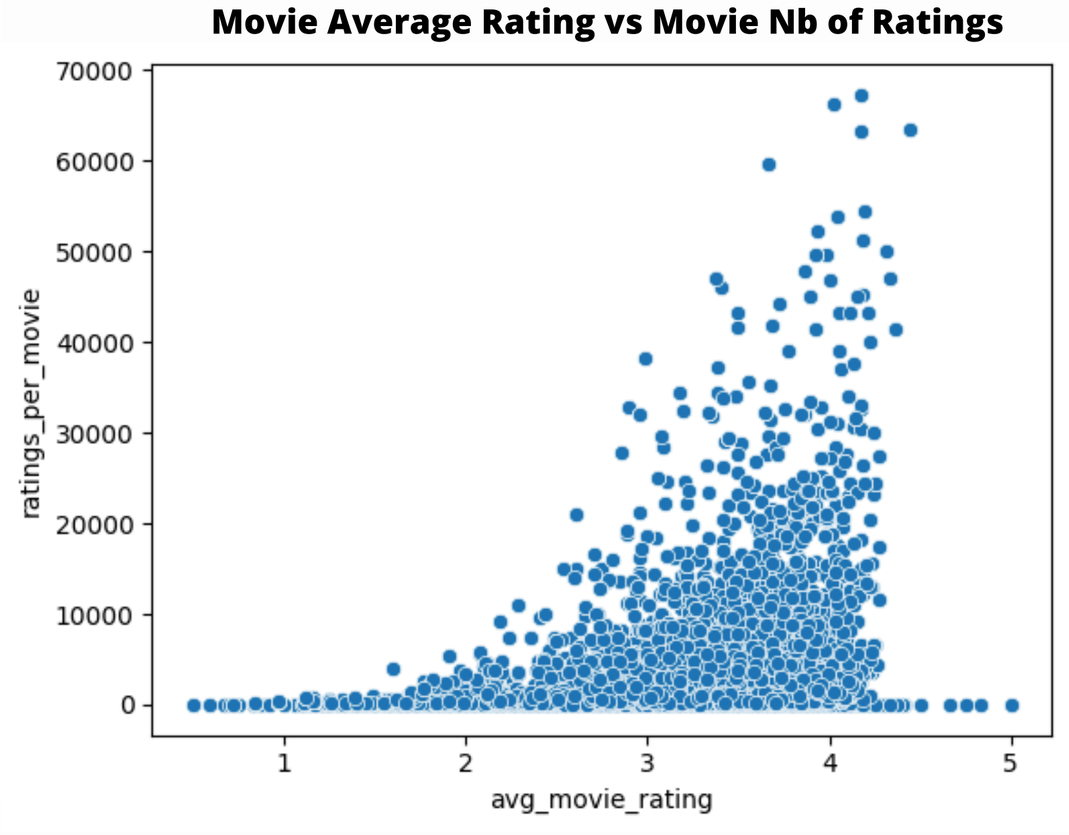
Users


Distribution of Ratings per User
Frequency
Number of Ratings per User
0
2000
4000
6000
8000
10000
12000
14000
16000
0
250
500
750
1000
1250
1500
1750
2000


User Average Rating vs User Nb of Ratings
ratings_per_user
avg_user_rating
0
2000
4000
6000
8000
1
2
3
4
5

Movie


Distribution of Ratings per Movie
Frequency
25000
20000
15000
10000
5000
0
0
10000
20000
30000
40000
50000
60000
70000
Number of Ratings per Movie


Movie Average Rating vs Movie Nb of Ratings
ratings_per_movie
70000
60000
50000
40000
30000
20000
10000
0
1
2
3
4
5
avg_movie_rating

DATA FILTERING

Top Active Users

- 644 -> 9254 ratings per user
- 5 millions ratings
- 25% of all ratings

- 26k movies
- 96% of all movies

3%

Middle Active Users

- 124 -> 643 ratings per user
- 10 millions ratings
- 50% of all ratings

- 20k movies
- 74% of all movies

28%

Less Active Users

- 0 -> 123 ratings per user
- 5 millions ratings
- 25% of all ratings

- 15k movies
- 57% of all movies

69%

Models

MATRIX FACTORIZATION

- PEARSON CORRELATION
- CORRELATION WITH OTHER MOVIES = COLLABORATIVE
- CORRELATION WITH GENRES = CONTENT-BASED
- HYBRID COLLABORATIVE AND CONTENT-BASED

SVD

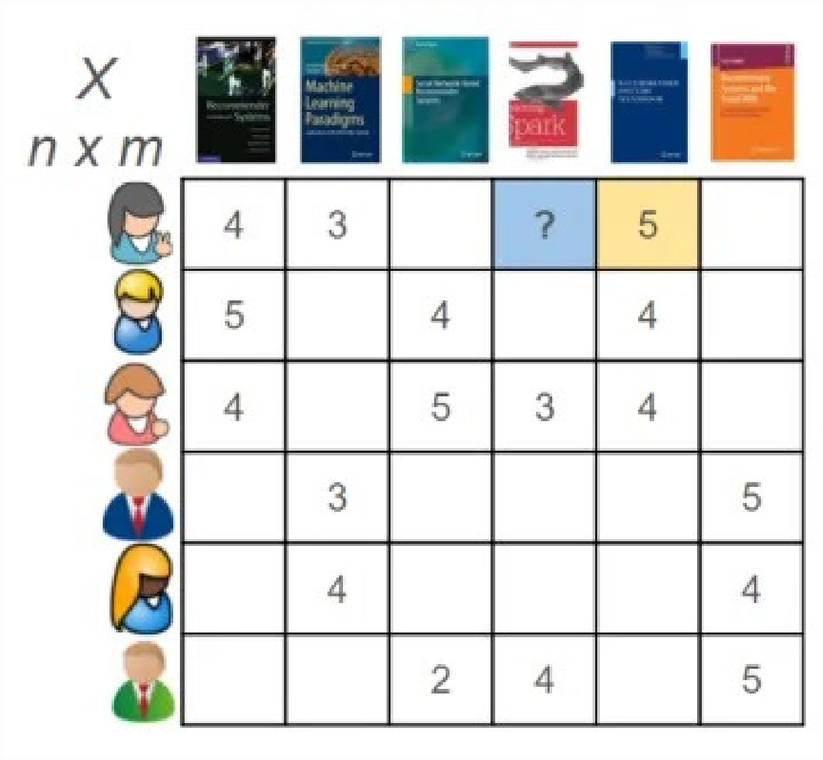
- ML MODELS BENCHMARK
- SVD ON FILTERED DATASET
- HYPERPARAMETER TUNING

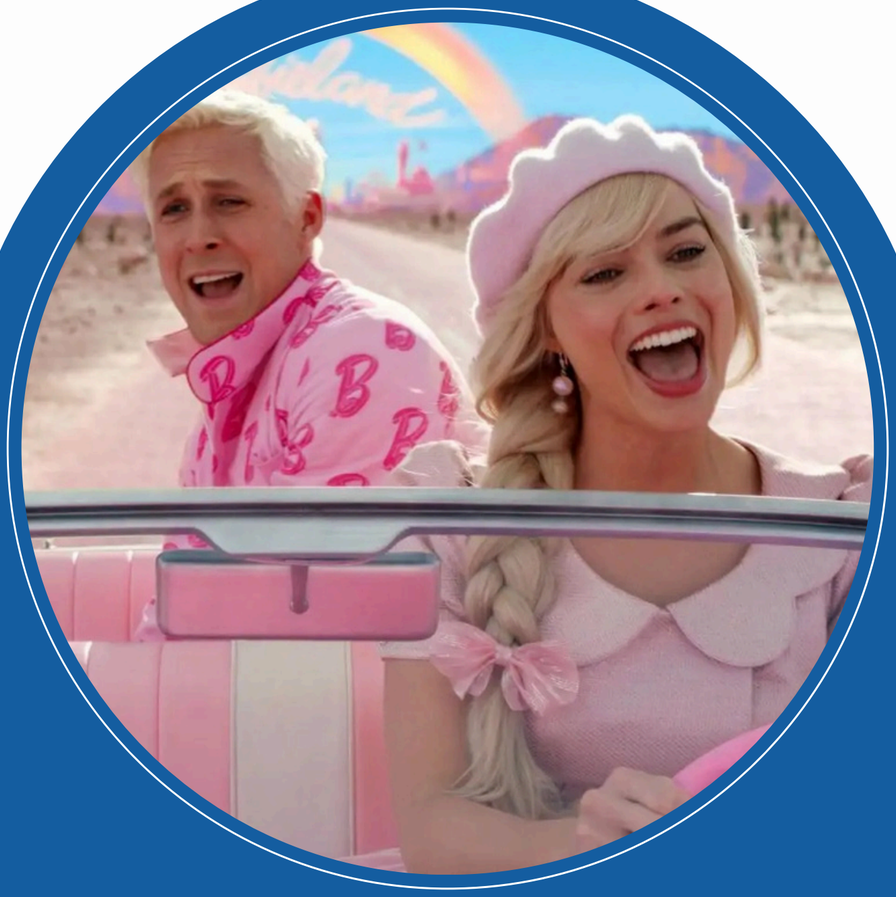
LightFM

- LIGHTFM EXPLAINED
- HYPEREPARAMETER TUNING
- MODEL EVALUATION

Basic Factorization Matrix

- Collaborative Recommendation
- Pearson Correlation





Adding Genre Correlation

- One Hot Encoding
- Content-Based Recommendation
- Pearson Correlation

Summary

- Because you watch...
- Simple and Efficient
- Popularity resistant

Machine Learning

Benchmark

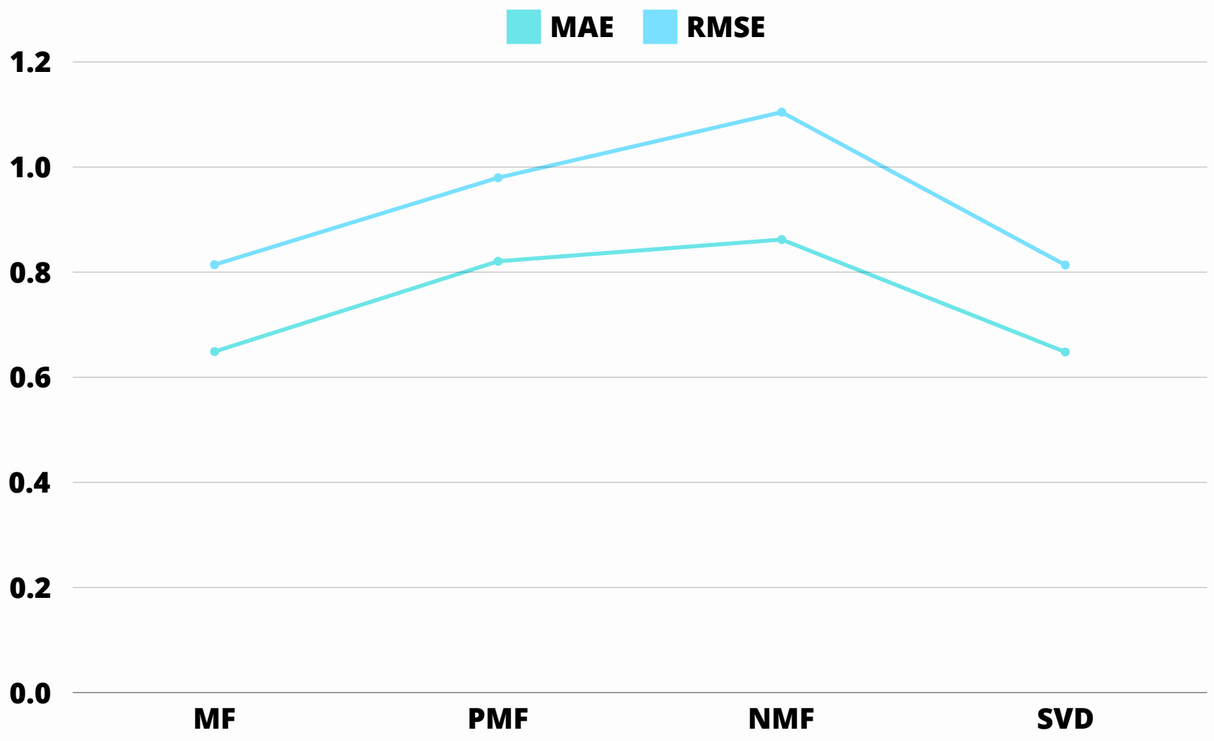
KNN (K-Nearest Neighbors)

MF (Matrix Factorization)

PMF (Factorisation Matricelle Négative)

NMF (Non-negative MatrixFactorization)

SVD (Singular Value Decomposition)



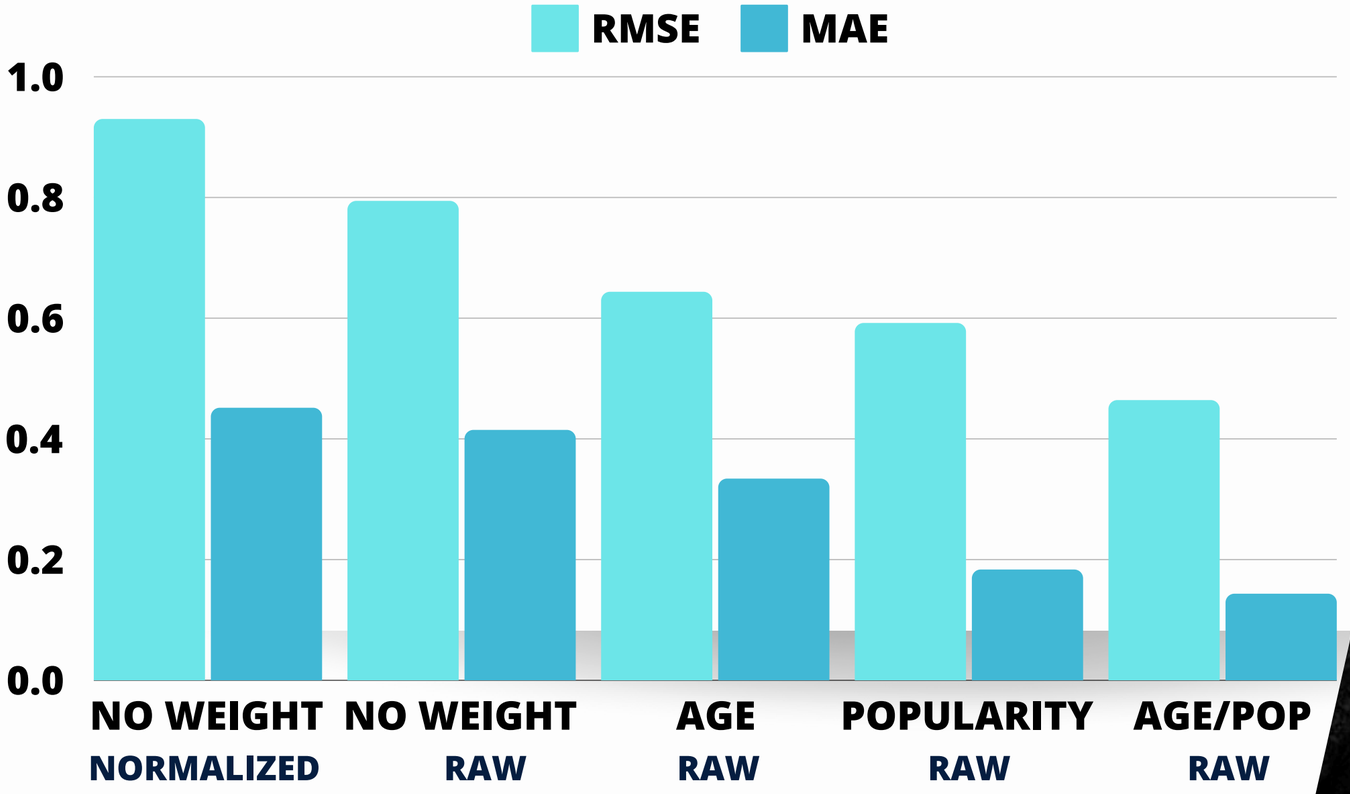
SVD

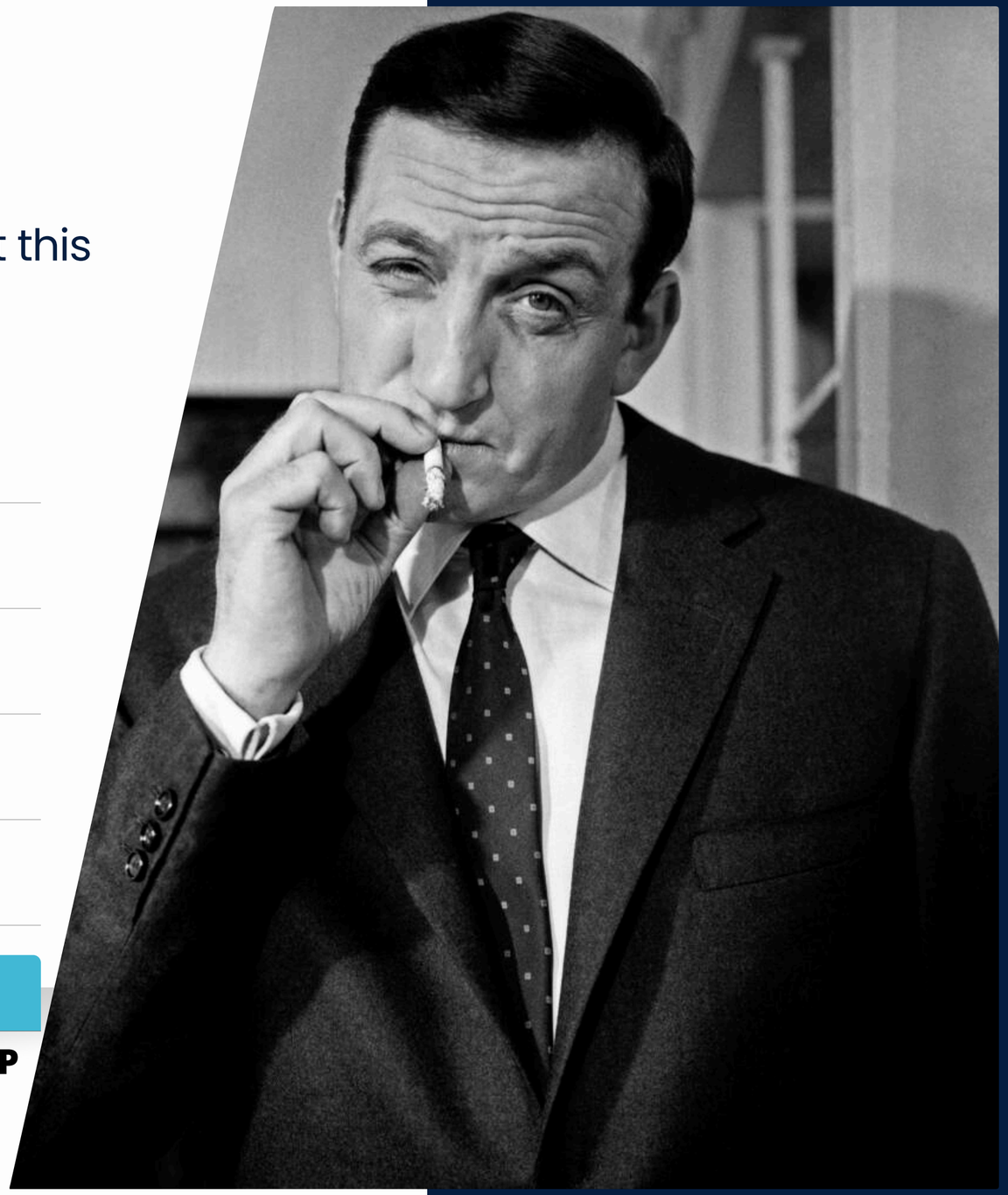
TARGET

- MAE : 0.585
- RMSE : 0.734

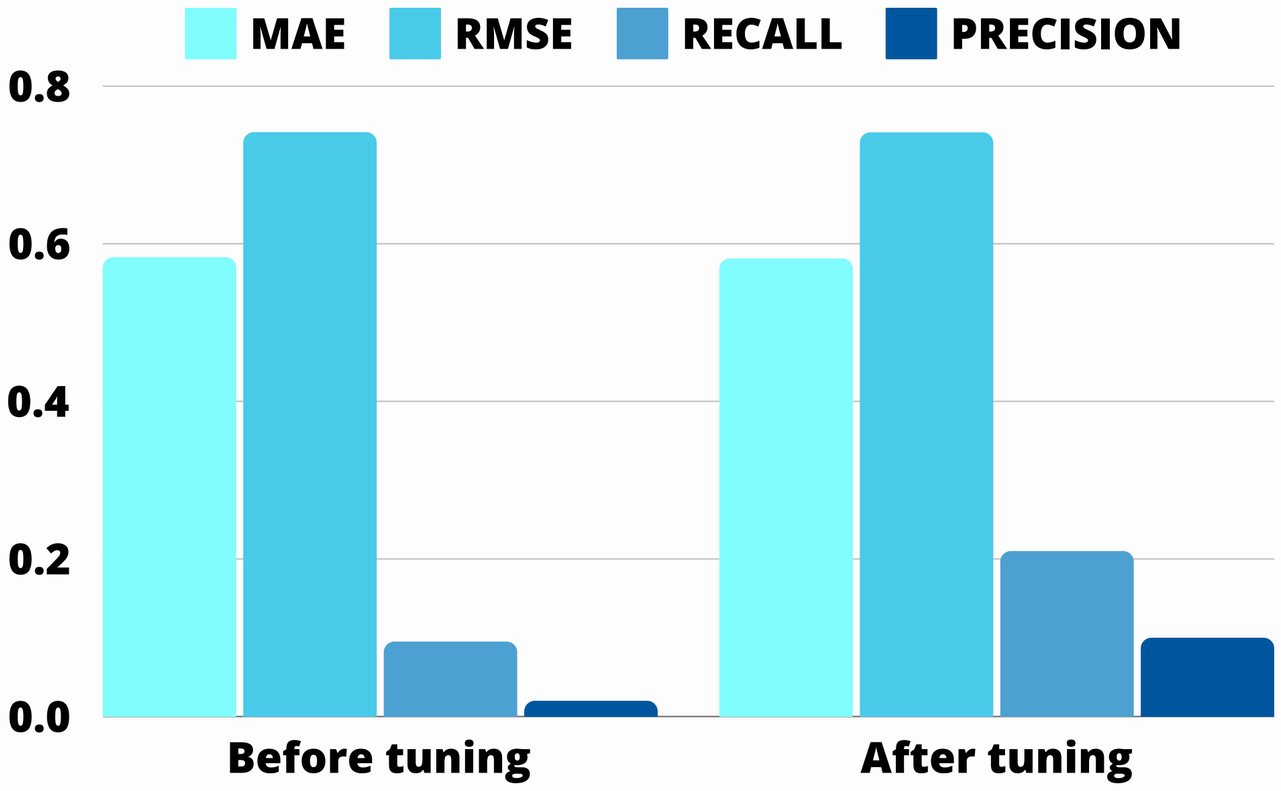
Using all panel of viewer

The first test with SVD, all tags with weighting get this result :





Using with IQR on dataset & Tuning



Test to improve recall and precision score

Recommendation user 1

	title	genres	rating
0	Freaks	Crime|Drama|Horror	5.0
1	Lord of the Rings: The Fellowship of the Ring,...	Adventure|Fantasy	5.0
2	Lord of the Rings: The Two Towers, The	Adventure|Fantasy	5.0
3	Lord of the Rings: The Return of the King, The	Action|Adventure|Drama|Fantasy	5.0
4	Star Wars: Episode V - The Empire Strikes Back	Action|Adventure|Sci-Fi	4.5
5	Raiders of the Lost Ark (Indiana Jones and the...	Action|Adventure	4.5
6	Spider-Man 2	Action|Adventure|Sci-Fi|IMAX	4.5
7	Constantine	Action|Fantasy|Horror|Thriller	4.0
8	Watership Down	Adventure|Animation|Children|Drama|Fantasy	4.0
9	Dragonslayer	Action|Adventure|Fantasy	4.0

```
Les recommandations faites sont les suivantes :
```

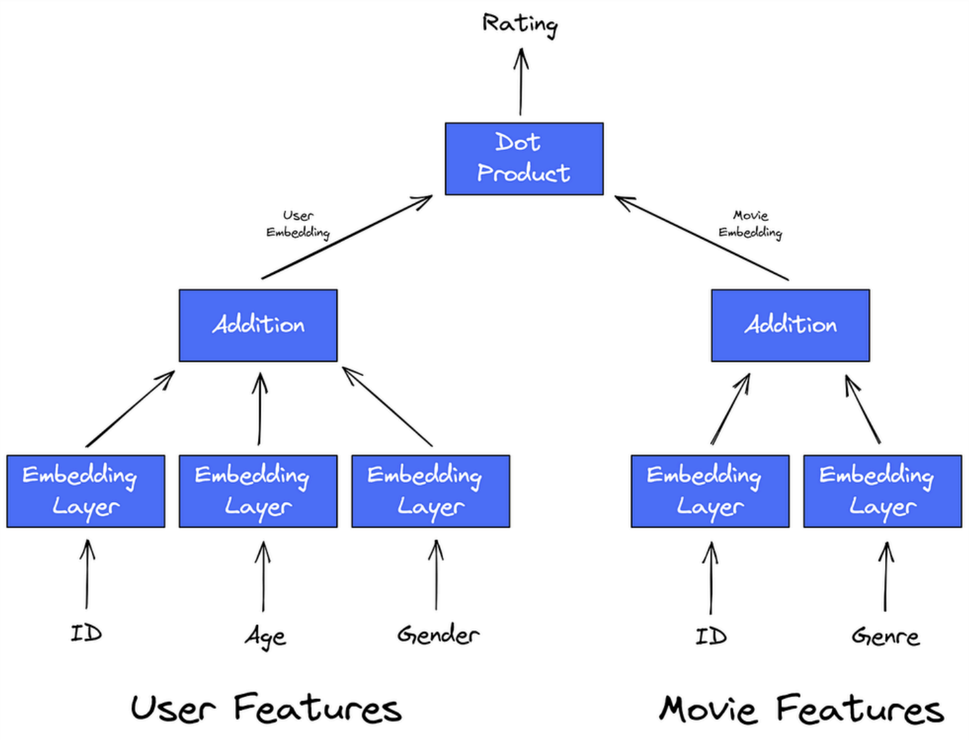
	title	genres	score
1	Jaws	Action|Horror	5.280891
7	Snatch	Comedy|Crime|Thriller	5.241471
4	Pi	Drama|Sci-Fi|Thriller	5.185878
8	Before the Devil Knows You're Dead	Crime|Drama|Thriller	5.160434
5	Dirty Dozen, The	Action|Drama|War	5.089878
2	Truman Show, The	Comedy|Drama|Sci-Fi	4.994098
6	Being John Malkovich	Comedy|Drama|Fantasy	4.937565
3	Primary Colors	Comedy|Drama	4.933546
9	Chain Reaction	Horror	4.904399
0	Wizard of Oz, The	Adventure|Children|Fantasy|Musical	4.873779

Summary

- Good recommendation based on genre
- No blockbuster
- Predict score greater than 5

LightFM

- Hybrid Recommendation
- Combines Collaborative and Content-Based Filtering
- Works with Implicit & Explicit Feedback



Model Data

Interactions

User-Movie Interactions
Rating >=4

10M

Movies

One Hot Encoded Genres

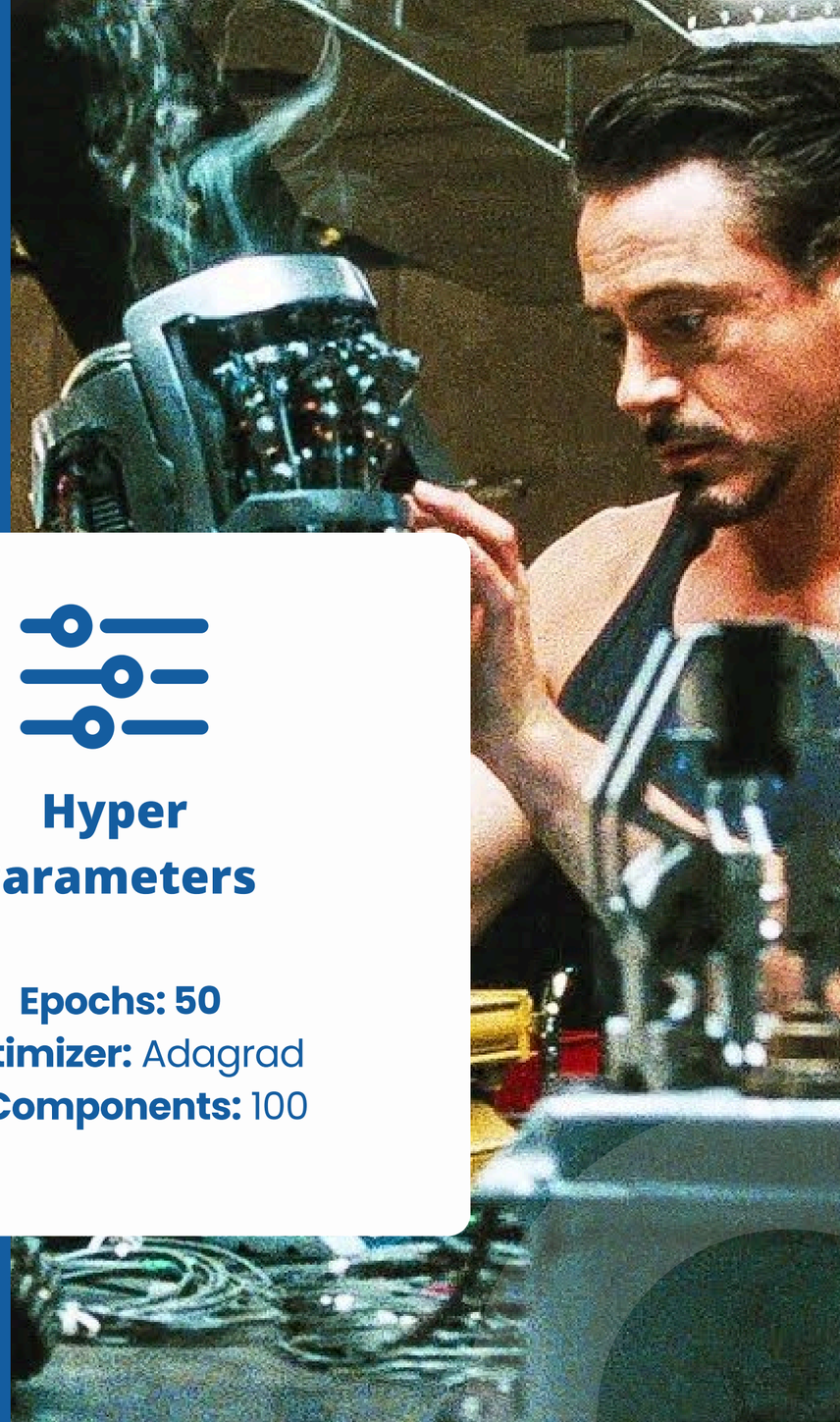
19 Genres

Users

Watched Genres
Distribution

38K Users

Model Optimization



Loss Function

Logistic
BPR
Warp
k-OS WARP

Dataset Split

30% Train

70% Test

Hyper Parameters

Epochs: 50
Optimizer: Adagrad
NoComponents: 100

Model Evaluation

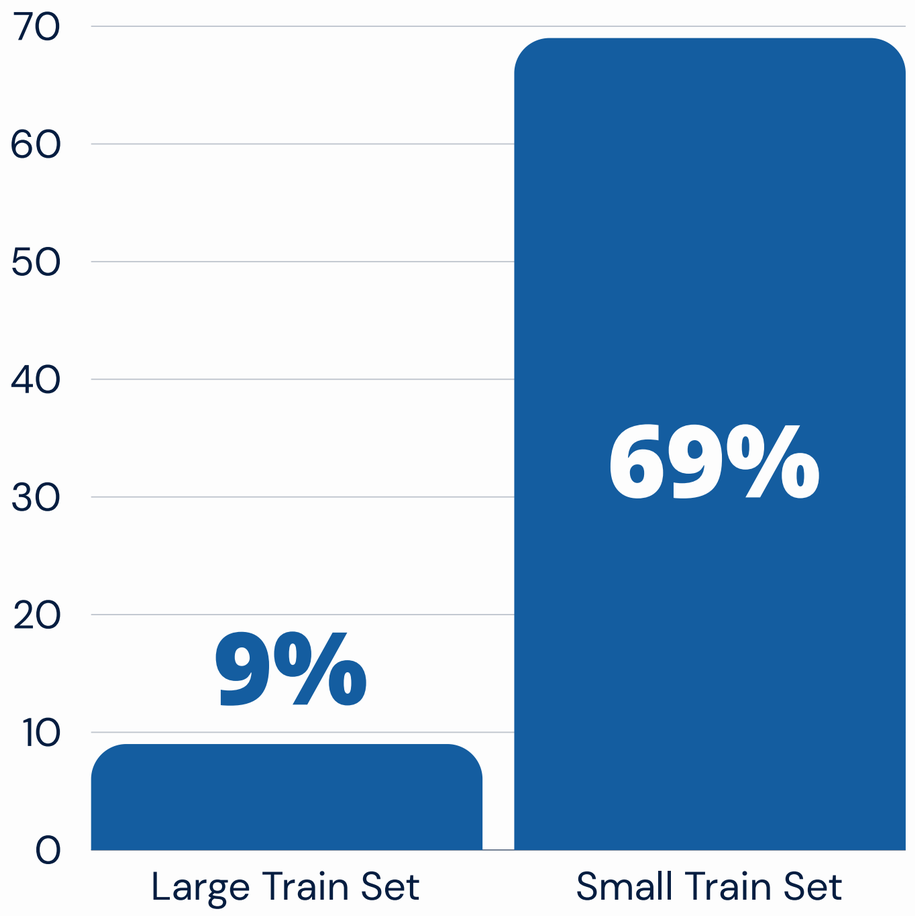
5 Folds
Cross Validation

69%
Precision

15
K-value

96%
Roc-AUC

Precision @15



Recommendation user 1

Watched Movies

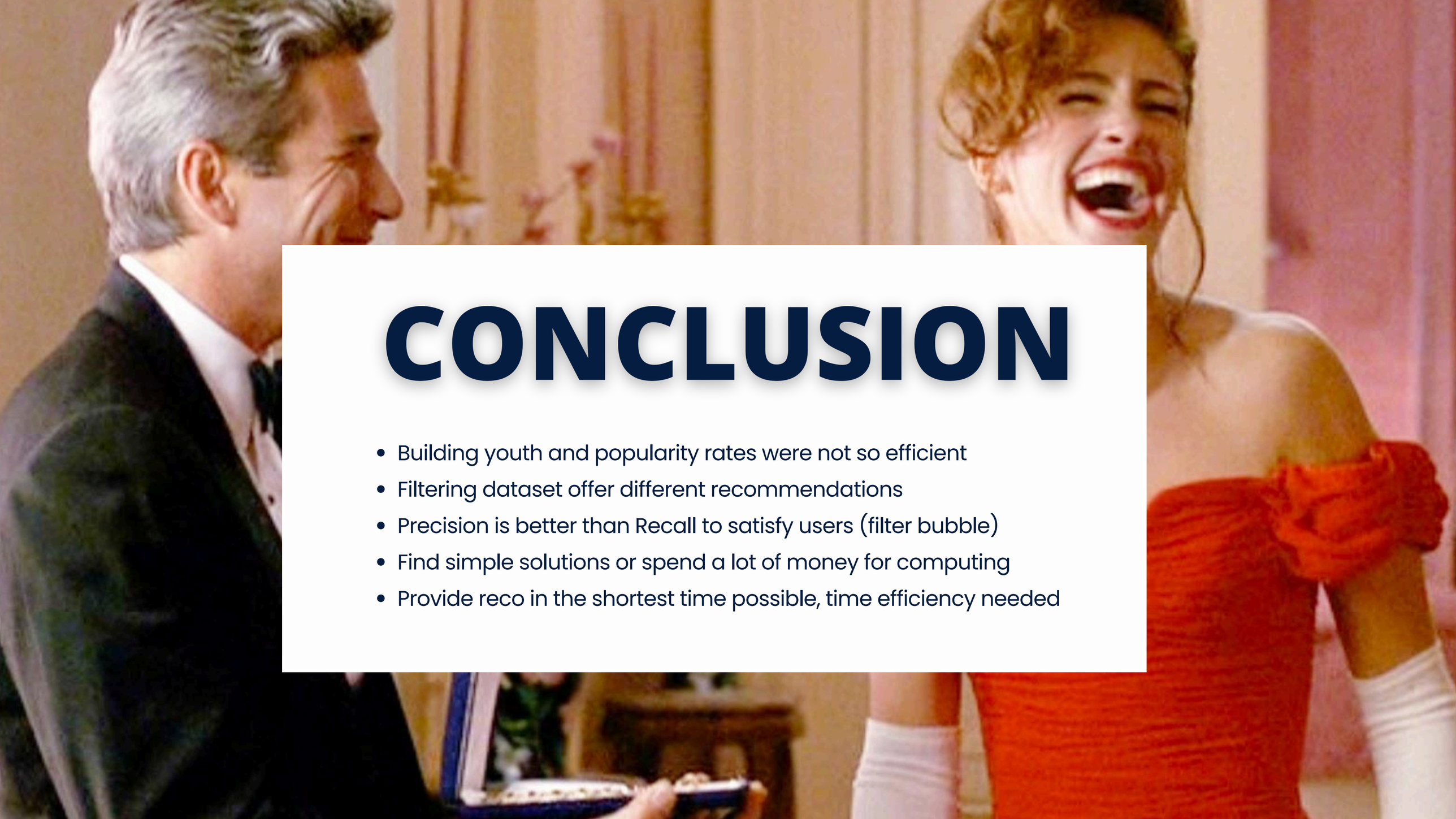
title	genres
Constantine (2005)	Action|Fantasy|Horror|Thriller
Splash (1984)	Comedy|Fantasy|Romance
Thing, The (1982)	Action|Horror|Sci-Fi|Thriller
Time Machine, The (2002)	Action|Adventure|Sci-Fi
Clash of the Titans (1981)	Action|Adventure|Fantasy|Romance
Willow (1988)	Action|Adventure|Fantasy
Beetlejuice (1988)	Comedy|Fantasy
Navigator: A Mediaeval Odyssey, The (1988)	Adventure|Fantasy|Sci-Fi
Legend (1985)	Adventure|Fantasy|Romance
Dark Crystal, The (1982)	Adventure|Fantasy

Recommended Movies

movieId	title	genres
36401	Alien (1979)	Horror|Sci-Fi
2918	Psycho (1960)	Crime|Horror
2272	One True Thing (1998)	Drama
1214	You've Got Mail (1998)	Comedy|Romance
2687	Mummy, The (1932)	Horror|Romance
1219	Tarzan (1999)	Adventure|Animation|Children|Drama
2633	Ferris Bueller's Day Off (1986)	Comedy
5066	Rat Race (2001)	Comedy
2424	Walk to Remember, A (2002)	Drama|Romance
4728	Brothers Grimm, The (2005)	Comedy|Fantasy|Horror|Thriller

Summary

- Adapted Recommendation
- Good model for both frequent users & Cold-Warm start Cases



CONCLUSION

- Building youth and popularity rates were not so efficient
- Filtering dataset offer different recommendations
- Precision is better than Recall to satisfy users (filter bubble)
- Find simple solutions or spend a lot of money for computing
- Provide reco in the shortest time possible, time efficiency needed

Our Team

Mohamed Kallel

Simon-Pierre Rodner

Jean-Christophe Rigoni

THANK YOU!