

# Presentation

---

## Research The Techniques of Recommender System on Book Dataset

Master: Quách Đình Hoàng

Trần Nguyên Thái Bảo      19133010  
Lê Tuấn Hiệp      19133023  
Đình Quốc Hùng      19133025

---

# Goal of the project

- ❖ Research The Techniques of Recommender System on Book Dataset
- ❖ Help user to find the right books
- ❖ Output: ratings of book which user will rate
- ❖ Evaluation methods: Root-mean-square-error (RMSE)

## Dataset: Book-Crossing Dataset

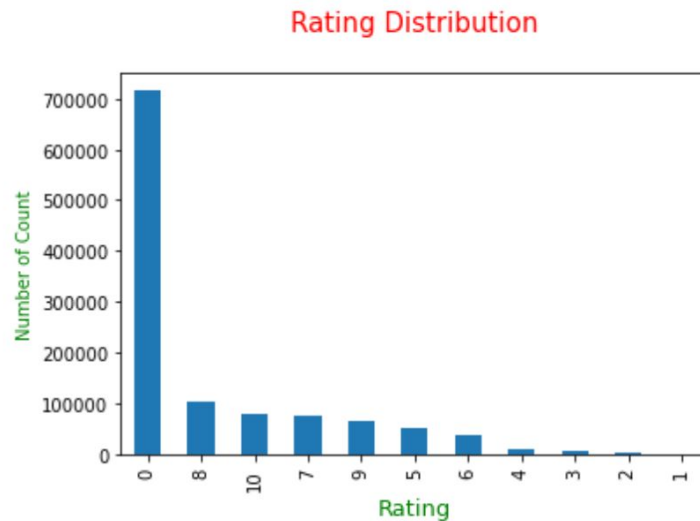
-Overview: Collected by Cai-Nicolas Ziegler in a 4-week crawl (August / September 2004) from the Book-Crossing community with kind permission from Ron Hornbaker, CTO of Humankind Systems. Contains 278,858 users (anonymized but with demographic information) providing 1,149,780 ratings (explicit / implicit) about 271,360 books.

-Format: 3 file csv (Book-rating.csv, Book.csv, User.csv)

-Link data: <http://www2.informatik.uni-freiburg.de/~ciegler/BX/>

# Solutions

- Using techniques of Recommendation system
- Using pyspark libraries
- **Content based:** using TF-IDF to create item profiles then predict ratings.
- **Collaborative filtering:**
  - + *Memory based:* Using pearson correlation metric (item-item, user-user)
  - + *Model based:* using ALS and SVD model



## Content based

- TF-IDF with title and using linear model to predict ratings.

---

```
RMSE for train: _ : 6.120163756754929  
RMSE for test   : 6.01752427252029
```

=> too bad result

# User-User collaborative filtering

```
[('273418', {'0671521438': 0.5, '1551665484': -0.5, 'row_avg': 6.5}),  
 ('259812', {'0446611212': 0.0, 'row_avg': 3.0}),  
 ('60083',  
  {'0739417096': 0.6666666666666661,  
   '0446519790': -0.3333333333333339,  
   '0505524643': -0.3333333333333339,  
   'row_avg': 8.333333333333334}),  
 ('273483',  
  {'1551667517': -0.6666666666666667,  
   '0446679593': 0.33333333333333304,  
   '0451193261': 0.33333333333333304,  
   'row_avg': 7.666666666666667}),  
 ('184513',
```

RMSE for user-user model: 3.2025806492887643

# Item-item collaborative filtering

```
[('0671521438', {'273418': -0.5, '153662': 0.5, 'row_avg': 7.5}),  
 ('0345391055',  
  {'267444': 0.5625,  
   '123544': -6.4375,  
   '27204': 1.5625,  
   '41841': -4.4375,  
   '61798': 1.5625,  
   '3965': 1.5625,  
   '14698': -0.4375,  
   '26339': 1.5625,  
   '234597': 1.5625,  
   '209373': 0.5625,  
   '44588': -2.4375,  
   '180942': -1.4375,  
   '41455': -2.4375,
```

RMSE for item-item model: 2.3952283538316923

# SVD

```
{'n_factors': 80, 'lr_all': 0.005, 'reg_all': 0.04}
```

```
RMSE: 1.6330
```



# ALS

```
**Best Model**  
Rank: 10  
MaxIter: 10  
RegParam: 0.1  
Alpha: 100.0
```

---

Root-mean-square error = 2.464188595062904

# Original Result

Technique	RMSE
SVD	3.5546
User-Based	4.3772
Item-Based	4.3435
ALS	4.6072

# Filter Data

- Remove ratings = 0
- Add column id for books

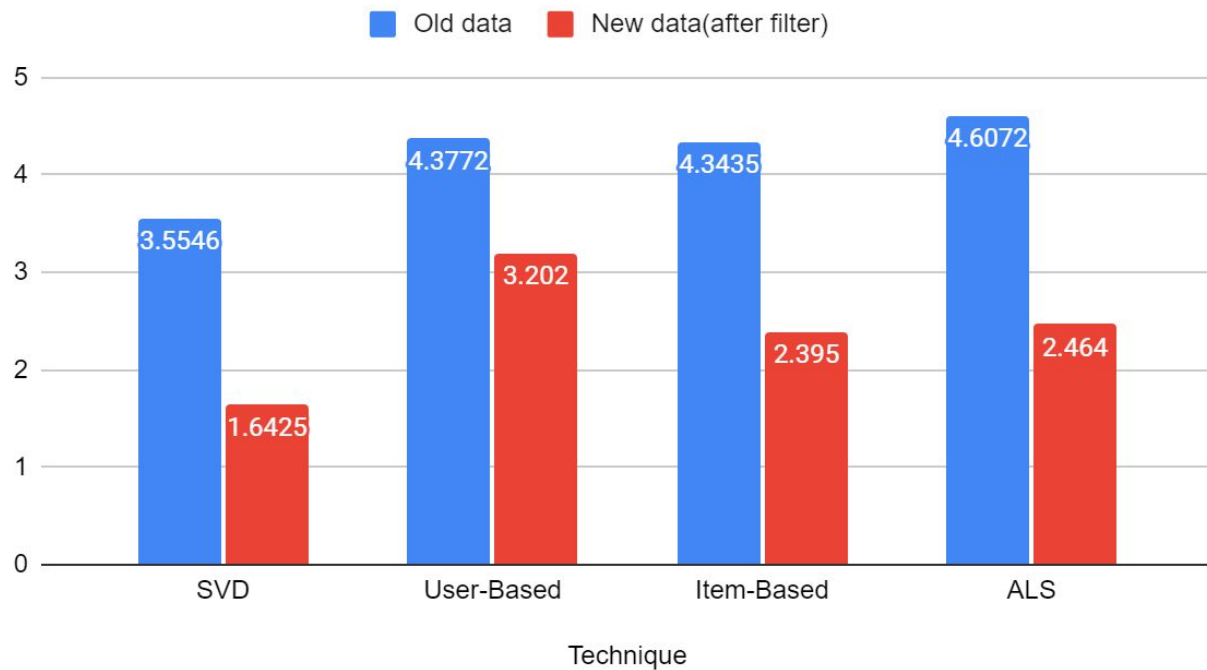
```
shape: (383842, 3)
```

## Results

Technique	RMSE	RMSE (filter data)
SVD	3.5546	1.633
User-Based	4.3772	3.202
Item-Based	4.3435	2.395
ALS	4.6072	2.464

# Statistical chart

Compare result



---

---

**THANK YOU FOR  
LISTENING**

---

---