Book Ratings

Giving the best book recommendations in a nutshell.

Book Ratings

01 Explaining the business problem

02 Data Exploration and Preparation

03 Modelling

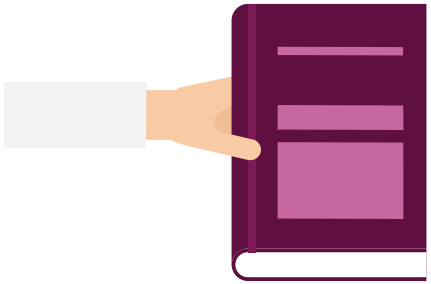
04 Interpretation

Business Problem

In a world in constant change and rapid growth, books have been losing their value. Therefore, it is important to apply strategies to increase their attractiveness (or, in a financial view, increase sales), to do that, a machine learning algorithm will be created to predict books you would actually might want to read.

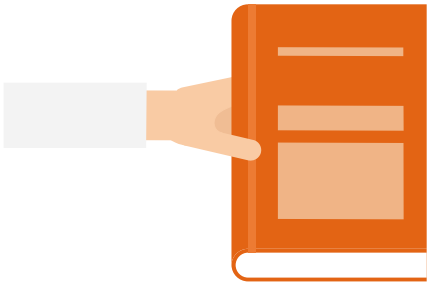
Goal: Building a ranked list of book ratings
Target: Unknown ratings

3 Datasets



Users

- User-ID
- Location
- Age

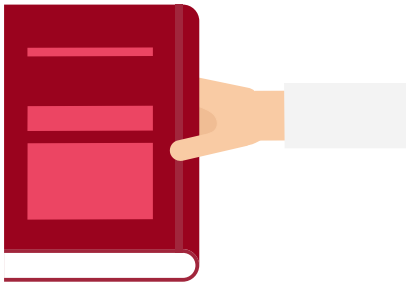


Ratings

- User-ID
- ISBN
- Book-Rating

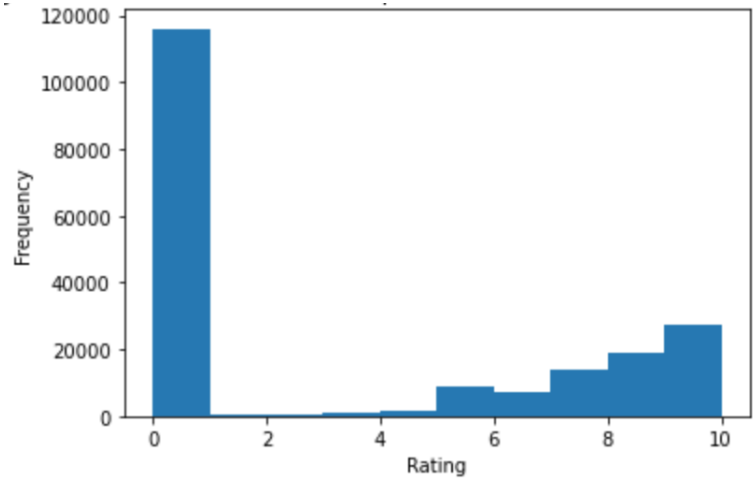
Books

- ISBN
- Title
- Author
- Year of publication
- Publisher
- Image URLs (S, M, L)



Book Ratings

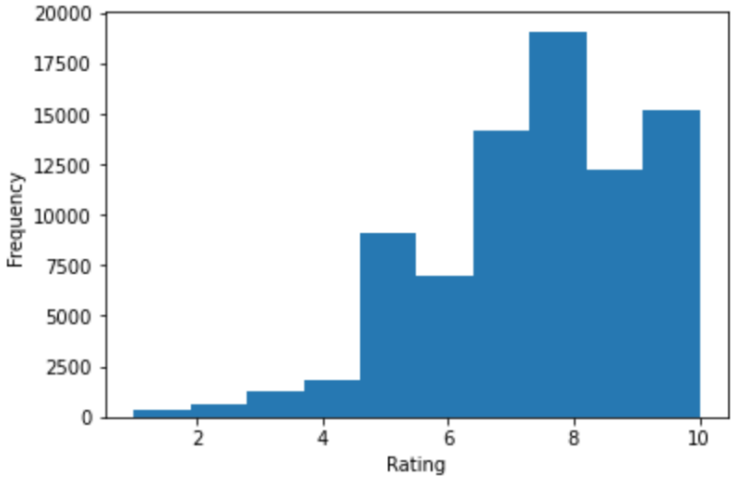
01 Overall Ratings distribution



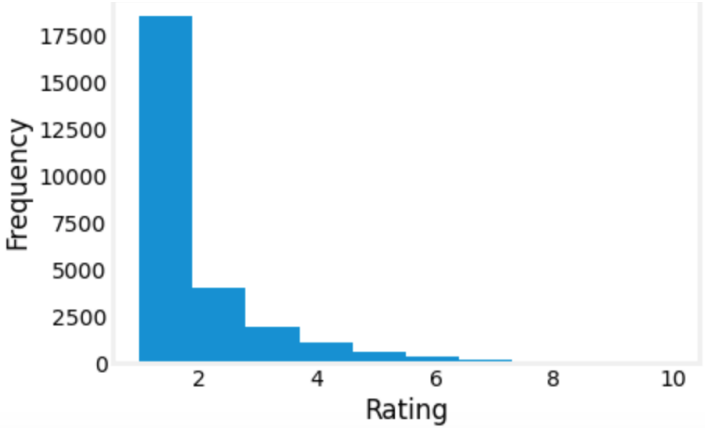
- 196,842 observations
- 116,036 zeros (users that bought books but gave no rating)
- Cannot use implicit and explicit ratings on modelling
- Drop the zeros and get 80,806 observations

Book Ratings

02 Ratings distribution without implicit ratings

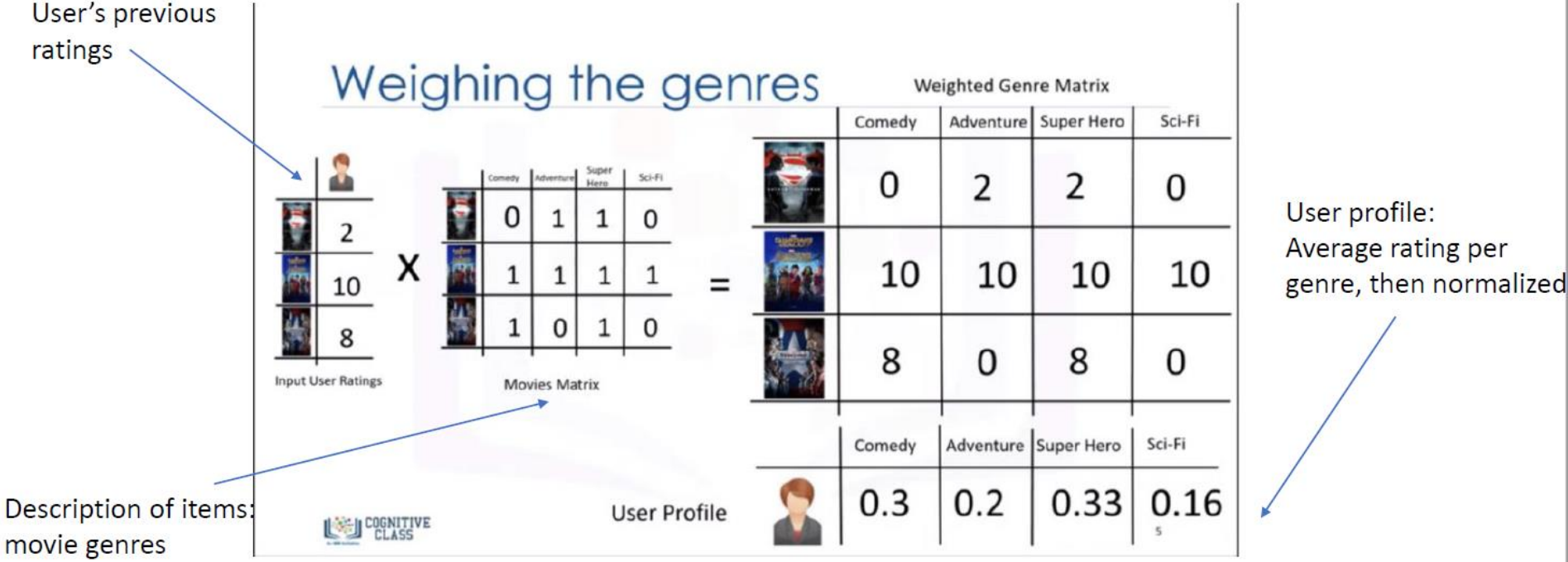


03 Ratings by user distribution



Modeling

Content-based Recommender System



Content-based recommender system - Problems

Computationally expensive algorithm, **1,076,509,530** predictions

A

3 ratings per user on average

B

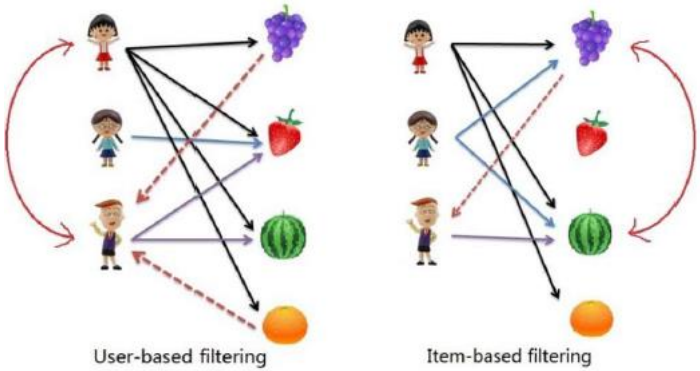
Lack of relevant features

C

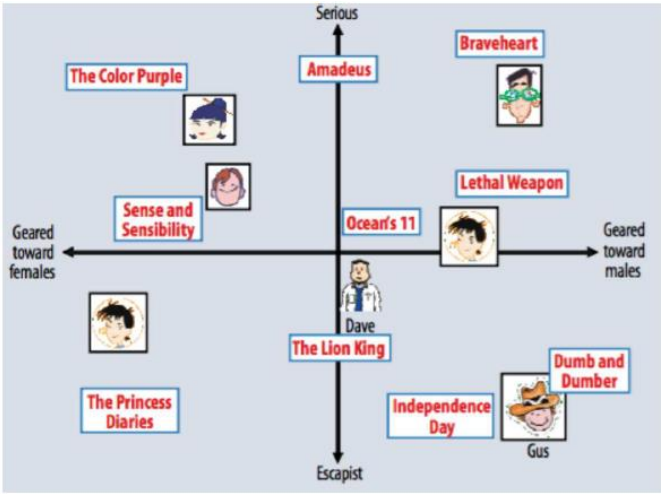
Modeling

Collaborative Filtering

Memory-based



Model-based



Results

01 Model-based Collaborative Filtering

02 RMSE

In-sample: 1.733

Out-of-sample: 1.707

03 Top 10 books

	Book	Predicted Rating
469	Harry Potter and the Order of the Phoenix (Boo...	8.73
393	Harry Potter and the Sorcerer's Stone (Harry P...	8.65
109	The Time Traveler's Wife	8.58
648	The Curious Incident of the Dog in the Night-T...	8.57
1094	Charlotte's Web	8.54
478	Name Der Rose	8.45
2324	Illuminati.	8.40
67	Jane Eyre	8.39
260	The Mount: A Novel	8.38
380	Sister of My Heart	8.37

Interpretability

What are the underlying factors? We don‘t know

To get an idea: Plot books / users in feature space,