Manga Recommender

Built with scikit-surprise and streamlit by Bryan Soh

Background

The Japanese Manga Market saw an **increase in sales by 10.3% in 2021** to reach 675.9 Billion Yen. This pushed manga sales to 40.4% of all domestic publishing sales in Japan, **crossing the 40% mark for the first time in history** (AnimeHunch).

Over in America, it was reported that manga sales made up **76.71% of all Adult Fiction Graphic Novel Sales in 2021** (AnimeNewsNetwork).

AnimeHunch

100,000+

Manga titles and growing!

京東

ホ
マ
ハ

OBJECTIVE

*Create a recommender system to reduce the barrier of entry
into the world of manga*

TABLE OF CONTENTS

01
Web Scraping
Scrape the web for user manga ratings

02
EDA
Clean and explore the scraped data

03
Modeling
Find the best performing machine learning model

04
App Deployment
Design and deploy the recommender system on the internet

01

Web Scraping

Where is the data from?

ホテル

東京

Data Source

MyAnimeList (MAL)
The world's largest online anime and manga community

User Clubs
User communities within MAL

Manga ratings of users
Scraped rating lists of 16K+ users for a total of 1.16M ratings

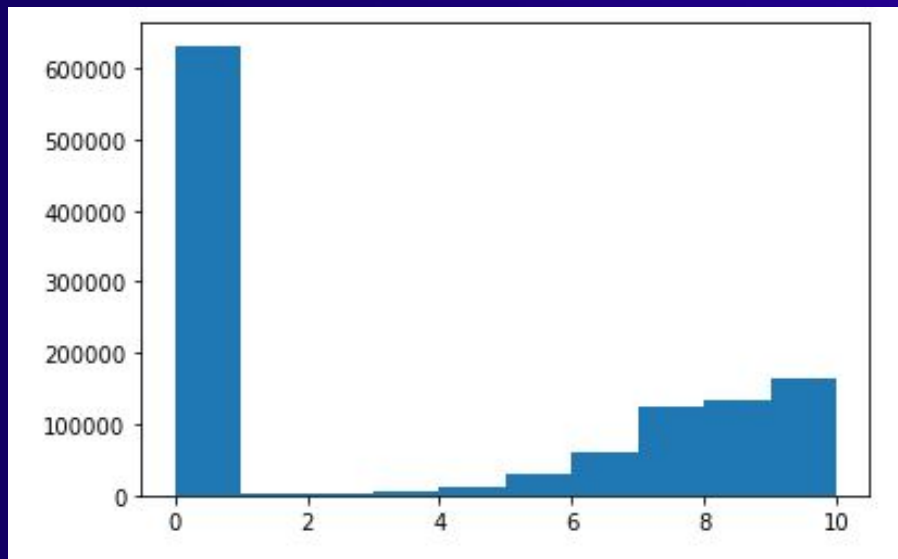**Recommendation Club**
The largest user club in MAL (~32K users)

02

EDA

Data exploration and manipulation

Disproportionate number of 0 ratings

54% of the 1.16M ratings were 0!

Top 20 manga titles

item	user	rating
Berserk	1503	1503
One Piece	1337	1337
Shingeki no Kyojin	970	970
Koe no Katachi	722	722
Oyasumi Punpun	714	714
Fullmetal Alchemist	700	700
Solo Leveling	663	663
Tokyo Ghoul	616	616
Death Note	608	608
Naruto	593	593
Vagabond	544	544
One Punch-Man	536	536
Chainsaw Man	440	440
Horimiya	422	422
Monster	400	400
Grand Blue	376	376
Fairy Tail	374	374
Vinland Saga	370	370
Bleach	354	354
JoJo no Kimyou na Bouken Part 7: Steel Ball Run	352	352

Highest number of 10 ratings

item	user	rating
Berserk	2747	2747
Oyasumi Punpun	2371	2371
Horimiya	2112	2112
One Punch-Man	1922	1922
Tokyo Ghoul	1887	1887
Shingeki no Kyojin	1881	1881
Vagabond	1872	1872
Boku no Hero Academia	1658	1658
Koe no Katachi	1621	1621
One Piece	1569	1569
20th Century Boys	1542	1542
Vinland Saga	1416	1416
Utsuro no Hako to Zero no Maria	1395	1395
Komi-san wa, Comyushou desu.	1372	1372
Solo Leveling	1344	1344
Kaguya-sama wa Kokurasetai: Tensai-tachi no Renai Zunousen	1313	1313
Grand Blue	1292	1292
Deadman Wonderland	1280	1280
Pandora Hearts	1253	1253
Monster	1213	1213

Highest number of 0 ratings

13/20 manga titles found in both lists

item	user	rating
Berserk	503	1503
One Piece	337	1337
Shingeki no Kyojin	970	970
Koe no Katachi	722	722
Oyasumi Punpun	714	714
Fullmetal Alchemist	700	700
Solo Leveling	663	663
Tokyo Ghoul	616	616
Death Note	608	608
Naruto	593	593
Vagabond	544	544
One Punch-Man	536	536
Chainsaw Man	440	440
Horimiya	422	422
Monster	400	400
Grand Blue	376	376
Fairy Tail	374	374
Vinland Saga	370	370
Bleach	354	354
JoJo no Kimyou na Bouken Part 7: Steel Ball Run	352	352

Highest number of 10 ratings

item	user	rating
Berserk	2747	2747
Oyasumi Punpun	2371	2371
Horimiya	2112	2112
One Punch-Man	1922	1922
Tokyo Ghoul	1887	1887
Shingeki no Kyojin	1881	1881
Vagabond	1872	1872
Boku no Hero Academia	1658	1658
Koe no Katachi	1621	1621
One Piece	1569	1569
20th Century Boys	1542	1542
Vinland Saga	1416	1416
Utsuro no Hako to Zero no Maria	1395	1395
Komi-san wa, Comyushou desu.	1372	1372
Solo Leveling	1344	1344
Kaguya-sama wa Kokurasetai: Tensai-tachi no Renai Zunousen	1313	1313
Grand Blue	1292	1292
Deadman Wonderland	1280	1280
Pandora Hearts	1253	1253
Monster	1213	1213

Highest number of 0 ratings

Removing all 0 ratings

2,200 users with only 0 ratings - Haters
7,194 titles with only 0 ratings - Nobody likes them

Number	Before	After	% Change
Users	16,463	14,263	-13%
Titles	32,363	25,169	-22%

Removing titles with few data points

Most popular title has 3,309 ratings
Remove titles with less than 1% of that (34 ratings)

Number	Before	After	% Change
Users	14,263	14,138	-0.9%
Titles	25,169	2,907	-88%

03

Modeling

Evaluating the best model

Evaluation Metrics

Precision@k and Recall@k

k= top 10 recommendations

$$P = \frac{\text{\# of our recommendations that are relevant}}{\text{\# of items we recommended}}$$

Precision

$$r = \frac{\text{\# of our recommendations that are relevant}}{\text{\# of all the possible relevant items}}$$

Recall

Relevance threshold set to >=7

Evaluation Metrics

A manga title with a rating of >=7 is considered relevant

Precision@k Score

In the recommender's top 10 recommendations, how many were relevant to the user?

Recall@k Score

Out of all relevant manga titles a user rated, how many were in the top 10 recommendations?

Baseline Model

Randomly pick 10 titles for each user

Algorithm Name	Precision@k	Recall@k
Baseline (Random Recommendations)	0.007795	0.003381

Calculate scores based on these recommendations

Evaluation Results

Algorithm Name	Average Precision@k Score	Average Recall@k Score
Basline Only	0.909319	0.862302
KNN Basic	0.901041	0.854036
KNN Baseline_Cosine	0.905999	0.852399
KNN Baseline	0.908539	0.850089
KNN Basic_Cosine	0.895532	0.848228
SVD	0.909108	0.845995
KNN with Z-Score_Cosine	0.894589	0.827692
KNN with Z-Score	0.896743	0.826443
KNN with Means_Cosine	0.890518	0.823206
KNN with Means	0.892933	0.823171
Co-clustering	0.893439	0.820467
Slope One	0.898290	0.815146
Normal Predictor	0.800907	0.646443
Non-negative Matrix Factorization	0.407405	0.206775
Baseline (Random Recommendations)	*0.007795*	*0.003381*

- *Scores shown are the average taken over 5 cross validation folds*

- *ML models vastly outperform the random experiments*

- *Baseline Only model chosen for best performance*

Baseline Only Model

$$\hat{r}_{ui} = b_{ui} = \mu + b_u + b_i$$

$\hat{r}_{ui}$ = Predicted Rating
b_{ui} = Baseline Estimate
μ = Average of all ratings
b_u = Bias of user u = Observed deviation of user u from average
b_i = Bias of item i = Observed deviation of item i from average

$$\min_{b_*} \sum_{(u,i)\in\mathcal{K}} (r_{ui} - \mu - b_u - b_i)^2 + \lambda_1 \left(\sum_u b_u^2 + \sum_i b_i^2 \right)$$

Biases are calculated using Alternating Least Squares (ALS) by default on scikit-surprise

04

App Deployment

Design and deploy the app on the internet

Streamlit App

This is my manga recommender

By Bryan Soh GitHub

Source code | Model from Scikit-Surprise

☐ You are above the age of 18 and wish to see adult recommendations as well

Choose the genres you want to see in your recommendations

Action ✕   ⊗ ▾

Select any number of your favorite manga titles

Akumetsu/アクメツ ✕   ⊗ ▾

Recommend

1) Check box for if you want to see adult recommendations
   a) Affects genres available in step 2

2) Select genres you want to see in your recommendations.
   a) Only titles containing at least 1 of the genres will be shown in the recommendations

3) Select any number of your favorite manga titles
   a) You can also search using alternative titles or Japanese titles

4) Click recommend!

Link to Streamlit App

Iphone users are recommended to try on laptop instead

THANKS!

To all the amazing people who are a part of DSIF3