The Kintoun Anime Recommender System

James Skelton

Goals

- Produce a recommendation generating system to predict users ratings of previously unseen anime.
- Use Spark and an Amazon Web Services Cluster
- For example, if i were to enter in 'Dragon Ball Z' (my favorite anime), my recommender might predict that i would like other similar anime such as One Piece or Hunter x Hunter.

The Source & Data

- MyAnimeList.net is one of the largest, most popular English language review and information websites for Japanese animation.
- It is considered by many English speakers to be one of the best places to review and discuss these programs.
- The original dataset consisted of 31284030 x 11 rows, as well as two supplementary information datasets about the anime and users individually.

MyAnimeList

Hide Ads | Login | Sign Up

Anime   **Manga**   **Community**   **Industry**   **Watch**   **Read**   **Help**

All ▼ | Search Anime, Manga, and more... | 🔍

One Piece

Details   **Videos**   **Episodes**   **Reviews**   **Recommendations**   **Stats**   **Characters & Staff**   **News**   **Forum**   **Featured**   **Clubs**   **Pictures**   **More Info**

Reviews

Write a Review

Sort ⌄

More Reviews

Maxine (All reviews)   Feb 28, 2008

3783 people found this review helpful   344 of ? episodes seen

Overall Rating: 10

There are two responses I get, without fail, every time I try to get someone new to watch this show. "I don't like the art style" or "I've seen the dub - NO THANKS." I'm guilty of both of these myself. But if there's one thing I need to stress before even getting started on this review, it's that the 4Kids dub is NOT One Piece. For the love of god, PLEASE do not think it is. If you've suffered the misfortune of seeing some of the 4Kids episodes, just erase them from your mind and start fresh. They read more

permalink | report

holdenn (All reviews)   Sep 12, 2008

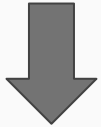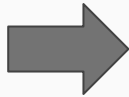**1460** people found this review helpful   379 of ? episodes seen

Overall Rating: 10

One Piece is by far the best shounen anime out there that I have watched.
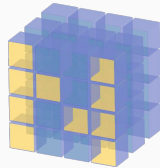But not all share the same views as I do, lets remedy that, shall we?
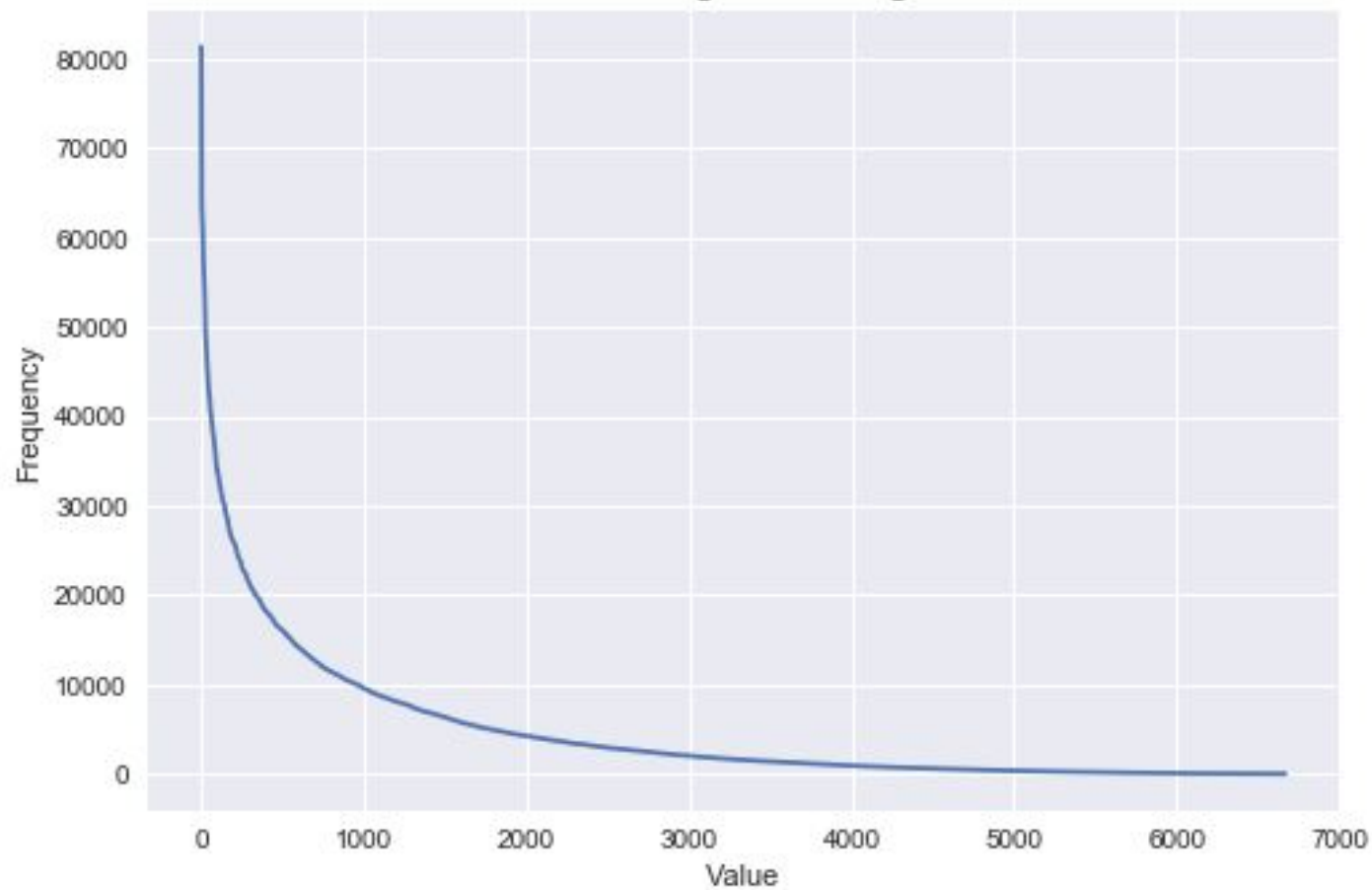
Workflow

Kaggle user Azathoth
scraped the data
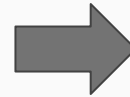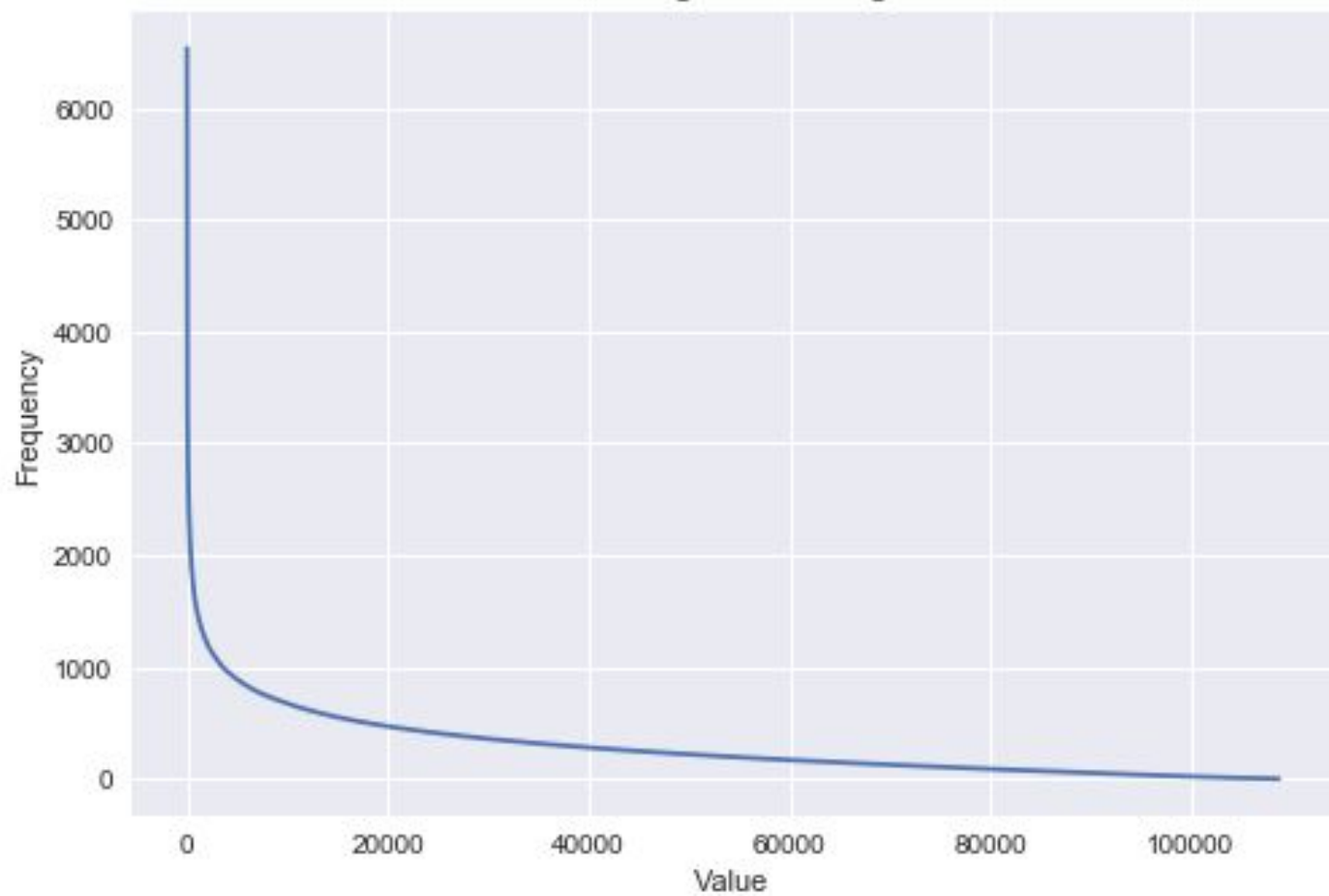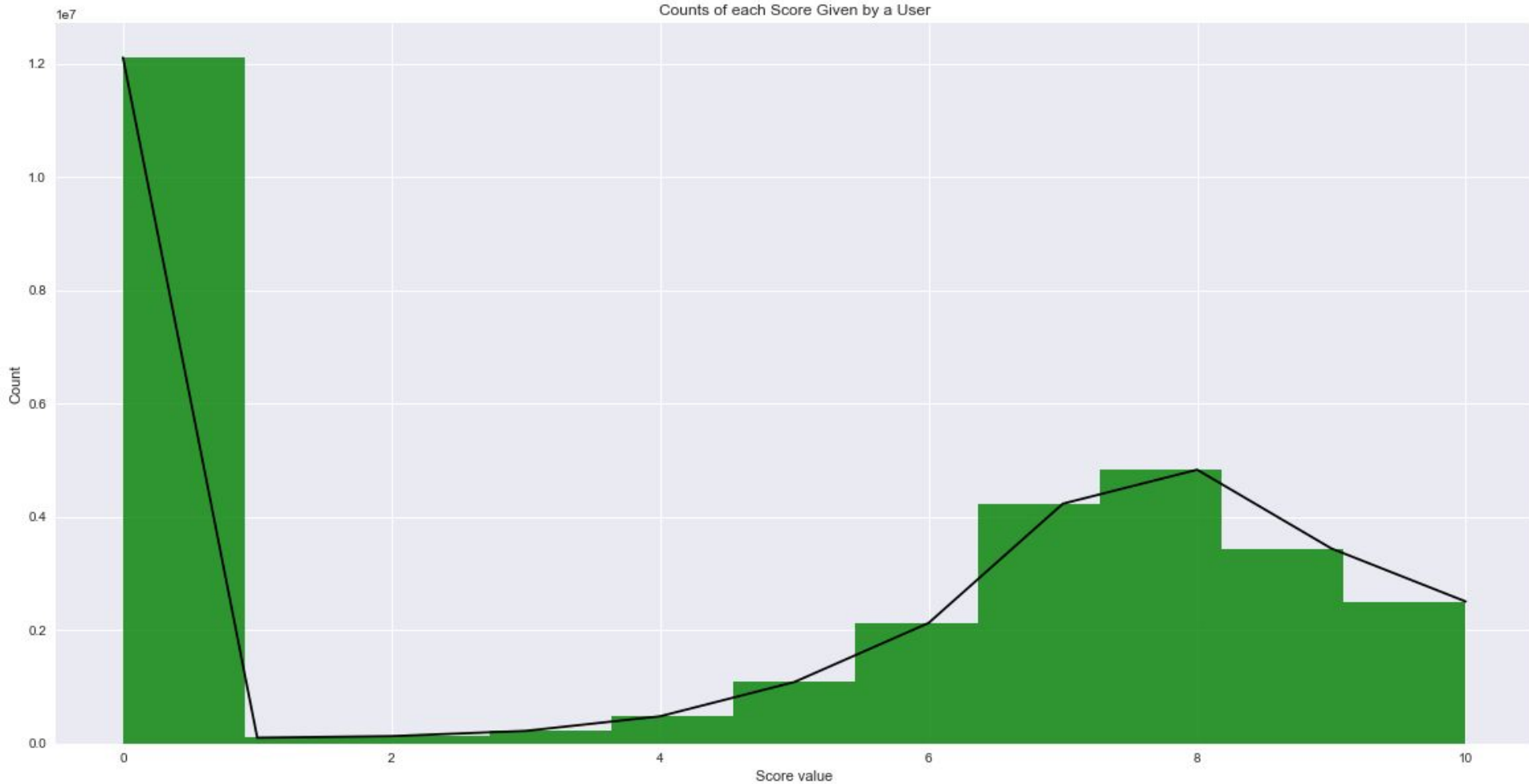
Anime Rating Count Histogram

User Rating Count Histogram

Counts of each Score Given by a User

Key Decisions

- 'My_status' > 2 were dropped.
- Any user that had put out less than 6 reviews was dropped due to insufficient data.
- Any anime with less than 6 reviews was also dropped due to insufficient data.
- When the recommender was unable to assign a value, the average rating for each anime was used to fill the missing values.

Model & Results

- Spark's Alternating Least Squares regression model was used for this analysis
- The parameters that yielded the lowest rmse were rank 220, maxIter 18, and regparam .1
- The Root Mean Squared Error was 1.614399
- I addressed the cold start problem by filling missing values with their average

Plot of the Predicted versus Sctual Scores
Rated on MyAnimeList
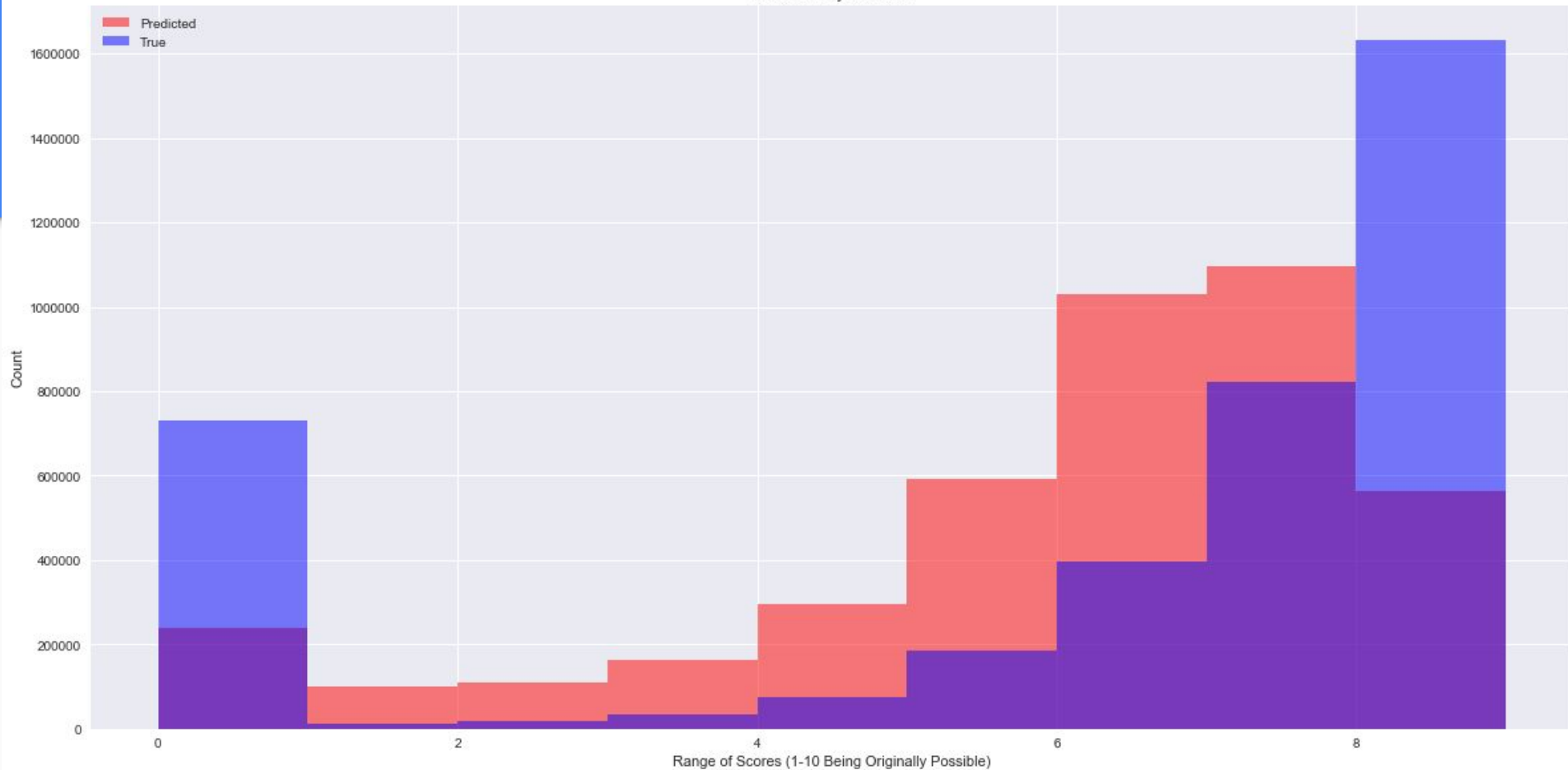
Next Steps

- There exists large, unused amounts of information about each anime and user in the original dataset. Incorporating these would offer a number of interesting options to counter further improve this model.
- Another model could be made using this same dataset and Natural Language Processing instead of matrix factorization

Thank you for listening

University of St. Andrews, BSc in Psychology, 2018

Project Github:
https://github.com/Jameshskelton/Anime_Recommender

Linkedin:
https://www.linkedin.com/in/james-skelton-b03aab1a1/

Contact email:
jameshskelton@gmail.com