Rolando Holmes

Derek Davenport

CSE 621-01: WEB MINING FOR E-CMRC

MAL (My Anime List) Sentiment Analysis Report

Objectives

For this assignment we proposed an idea to analyze the sentiment of user reviews for television shows. We focused exclusively on reviews of a narrow anime genre called “Isekai” . This term ‘Isekai’ is a Japanese word meaning “another world”. This genre consists of shows concerning a protagonist who is essentially whisked away into another world [**1**]. The concept of our project is analyzing user reviews for this genre from the **MAL (My Anime List)** website. MAL is a site where users write reviews, give scores, and interact with other users regarding anime shows. Although the reviewers do give a score to the show, we will only be conducting an unsupervised analysis of the text of the users’ reviews. We selected 14 well-known “Isekai” shows for analysis. We will scrape the MAL site for the user reviews on each of these shows, preprocess the user reviews to prepare the text for better analysis, and analyze the Polarity (Positive or Negative) sentiment and Subjectivity / Objectivity score for these shows.

Data Set Details

Our data from 14 shows should allow us to observe how they are perceived among each’s viewer base. How a show is reviewed will vary greatly from person to person. Some people share highly impassioned opinions from how a show made them feel, others point to concrete examples of successes or flaws. Going from review to review, it can be hard to get a sense of how a show’s audience perceives that show in general. Sentiment analysis can help us here.

Our data set has 783 samples in total, which can be seen in **Fig 1A**. It comprises of un-preprocessed user reviews. We also label each review for when we plot that Polarity and Subjectivity score, where we can designate a color to each label to distinguish them among the other plotted reviews. Below are a couple of examples of the *Not Processed* ***Fig.1*** and *Processed* data ***Fig.2.***

Fig.1*. Sample Dataset (Not Processed)*

	user_reviews	show
0	\n\n\n \n Re:Zero truly wanted to be something. It shoots for the stars, it tries new things and throws the characters through so much torture and misery that the fantasy-world they are living in resembles more a hell....	Re:Zero
1	\n\n\n \n *Re:minder - This review contains some light spoilers*\n\r\nFrom my perspective, we are living in a time where well crafted anime with a myriad of originality and detail are unfortunately being released at a ...	Re:Zero
2	\n\n\n \n "The only thing worse than dying once is having to die again."\n\r\nThat was the title of a wonderful email from Crunchyroll I received about a month and a half ago to advertise "Re: Zero - Starting Life in a...	Re:Zero
3	\n\n\n \n Hi all, this is my first review on MAL. Please bare with me, but I feel like I should share my opinion on this and why I gave it the score I did. I wouldn't say I'm the most hardcore anime fan, but I have def...	Re:Zero
4	\n\n\n \n Ahh, the reviewer. The first line of defense to quell the flames of irrational thought and often the bearer of bad news for those of the general public. They're the Buzz Killingtons that take it upon themselv...	Re:Zero

Fig.2*. Sample Dataset (Processed w/ Scores)*

	user_reviews	show	polarity	subjectivity
0	rezero truly wanted to be something it shoots for the stars it tries new things and throws the characters through so much torture and misery that the fantasyworld they are living in resembles more a hell they boil in this hell fuming with anger ...	Re:Zero	0.080578	0.574887
1	reminder this review contains some light spoilers from my perspective we are living in a time where well crafted anime with a myriad of originality and detail are unfortunately being released at a frequently declining rate because of this many i...	Re:Zero	0.190064	0.519002
2	the only thing worse than dying once is having to die again that was the title of a wonderful email from crunchyroll i received about a month and a half ago to advertise re zero starting life in another world i ignored it at first and simply car...	Re:Zero	0.145173	0.478223
3	hi all this is my first review on mal please bare with me but i feel like i should share my opinion on this and why i gave it the score i did i wouldnt say im the most hardcore anime fan but i have definitely seen my fair share of shows and am p...	Re:Zero	0.104430	0.524721
4	ahh the reviewer the first line of defense to quell the flames of irrational thought and often the bearer of bad news for those of the general public theyre the buzz killingtons that take it upon themselves to blow the no fun whistle and send th...	Re:Zero	0.062071	0.491544

Background Methodology

Web Scraping

We make use of the '**Beautiful Soup**' Package which allows us to scape webpages looking for specific elements of a webpage. In this case we are looking for div element that contain the all the user reviews for a given page. This process has an **$O(n^3)$** runtime with respect to the number of pages we are pulling from.

Preprocessing our data

Once we have gathered all our data, we sort it in a dictionary by show, this has an **$O(n)$** runtime with respect to length of the list of urls we are pulling our data from. We then take the data from the dictionary and build a DataFrame from it. The DataFrame has various filters applied to each review: a regex method to remove punctuation, stop-words removed, and nonsensical words removed. This has a **$O(2^m + n)$** run time.

Calculating Polarity and Subjectivity

The Polarity analysis is performed through the use a 'Naive Bayesian Analyzer" method that on average runs at **$O(nk)$** with respect to the number of features and classes. This method of classification sentiment was derived from the Stanford NLTK 'Naive Bayes" algorithm for analyzing a polarity score in text. A detailed explanation of model can be found here **[3].** We use a python package called 'Text Blob" **[4]**, this package makes use of a Machine Learning (ML) model using a 'Naive Bayes' classifier, the formula used is denoted as,

P(Text && Classification n) = P(Text | Classification n) P(Classification n) **[3]**

where the model is trained on classifying Negative or Positives movie reviews initially, in which our data is then passed into the model for it to be determined if the text is positive or negative and what the average polarity and sentiment score should be.

The Text blob package also makes use of a method called 'Pattern Analyzer" that utilizes a pattern library **[5]**, which contains a lexical database for English words (among others) and their 'Polarity and Subjectivity' hand-tagged scores annotated. **Fig.3** shows how different words are valued. The weight of negative to positive words are ranged (-1.0 to +1.0) and objective to subjective words are weighted (+0.0 to +1.0) **[6].** The final calculation is the average subjectivity score over all the words in the given text.

Fig.3 *Lexical database for English words*

```
<word form="amusing" wordnet_id="a-01344485" pos="JJ" sense="providing enjoyment" polarity="0.7" subjectivity="1.0" intensity="1.0" confidence="0.9" />
<word form="anger" wordnet_id="v-1785971" pos="VB" sense="make angry" polarity="-0.7" subjectivity="0.2" intensity="1.0" confidence="0.9" />
```

Calculating Feature Correlation

The Pearson correlation coefficient measures linear correlation between two variables. It normalizes covariance analysis to the range -1 to +1. A measure of -1 means that as variable one increases, variable two decreases perfectly while a measure of +1 means that variable two increases perfectly. A measure of 0 shows no correlation between the variables. Measures of +1 or -1 are unrealistic in real data with some amount of Gaussian distribution. In general, a measure greater than .5 or less than .5 indicates strong correlation.**[7]** The complexity of calculating a correlation matrix is the same as a covariance matrix which is **$O(nm^2)$** for n vectors with m features.

Building Word Clouds

The DataFrame used above was also used as the basis for the word clouds. However, some additional filtering was necessary to achieve the most useful results. Word clouds are meant to easily visualize the words that are used most frequently in a corpus. Since we are displaying a word cloud for each show, and each show is in the same genre, we expect significant overlap in the words used to review the various shows. It would not be particularly interesting to see the same words over and over in each word cloud. Rather, we want to see what makes each show's reviews special and unique from the others.

So we run through the frequency counts for each show and sum up how often each word is used. We look at the top 50 words for each show. Any words that appear in half or more of these top word lists are filtered out. These are words common to reviews of shows in our genre and while perhaps interesting in that sense, they are not helpful when comparing one show to another.

Flow Chart for processing and visualizing our data.

Scraping the MAL Website for user reviews.

Generating a Dictionary for our unprocessed reviews

Splitting up our Data in two individual user reviews lists.

Storing the lists of user reviews into a Data Frame to be cleaned.

After running a couple of rounds of cleaning we I.e.
- removing stop words,
- punctuation

We then calculate the Polarity and Objectivity Scores for each individual user review.

Compute Pearson's linear correlation coefficient matrix.
Build Linear Regression Model for features with highest correlation.
Plot highest correlated features with regression line for visualization.

With our updated DataFrame, we then plot the user reviews on a graph with the Label’s color legend listed as reference to which color of user reviews correlate to which show.

With this solution we should now be able to analyze, for a given show, the user Objectivity and Subjectivity score.

We then lastly produce a word cloud to visualize what are the most frequent words and topics discussed among the users in their reviews. Certain words were filtered to show the most unique and discriminatory terms to each show’s reviews.

Experimental Results

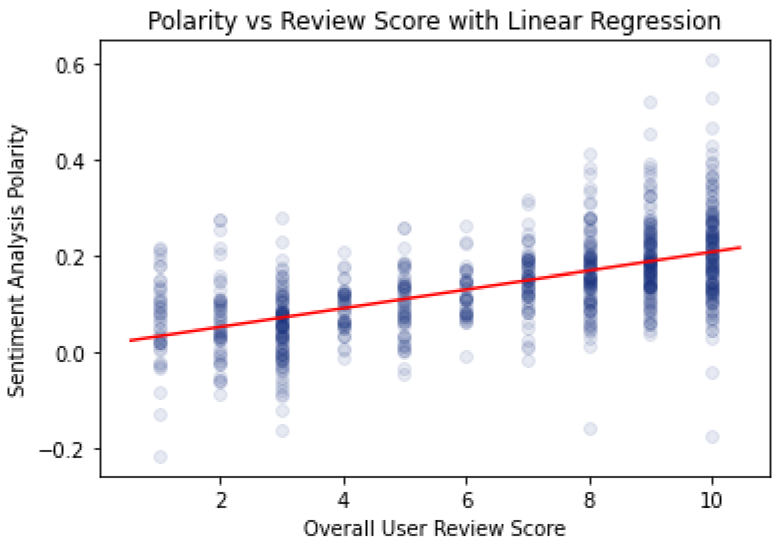
We did a Pearson correlation coefficient analysis across the entire dataset irrespective of show.

Fig 4 *Pearson Correlation Coefficient Matrix*

	overall	story	character	enjoyment	polarity	subjectivity
overall	1.000	0.812	0.837	0.837	0.578	-0.013
story	0.812	1.000	0.890	0.857	0.551	-0.004
character	0.837	0.890	1.000	0.876	0.549	-0.016
enjoyment	0.837	0.857	0.876	1.000	0.516	-0.022
polarity	0.578	0.551	0.549	0.516	1.000	0.083
subjectivity	-0.013	-0.004	-0.016	-0.022	0.083	1.000

In our dataset, we would expect the users' show review scores to highly correlate with the generated polarity score if the polarity score is a reliable metric. The Pearson correlation coefficient between overall score and polarity was computed to be 0.578 with a p-value extremely close to zero (~4e-71). As expected, there is a strong positive correlation. This confirms for us that we can have confidence that generated polarity score is a good measure of the text's positive or negative sentiment.

Fig.5



The plot in **Fig.5** shows the clear correlation. The correlation between overall review score and the other review scores was extremely high, but this is expected and uninteresting. The Pearson correlation coefficient between overall score and subjectivity was near zero. Since subjectivity is not trying to tell us how well a user liked a show, we would not expect much correlation here. Both positive and negative reviews can be based on facts as easily as they can be based on opinions. There was also not much correlation between polarity and subjectivity. Reviews can be equally positive and objective as positive and subjective or any other combination.

Below in **Fig.6** you can see an overview of the data used to plot the user sentiment reviews where in **Fig.7** you can also see the corresponding legend for what shows pertain to what show the user reviewed. We plot '**Polarity'** on the x axis and **'Subjectivity'** on the y axis. Although it is possible to identify some trends and groupings, we also created plots that show only one show's color at a time while all other shows are gray (see **Fig.8 and Fig.9).**

Fig.6 *Sentiment Analysis Plot*

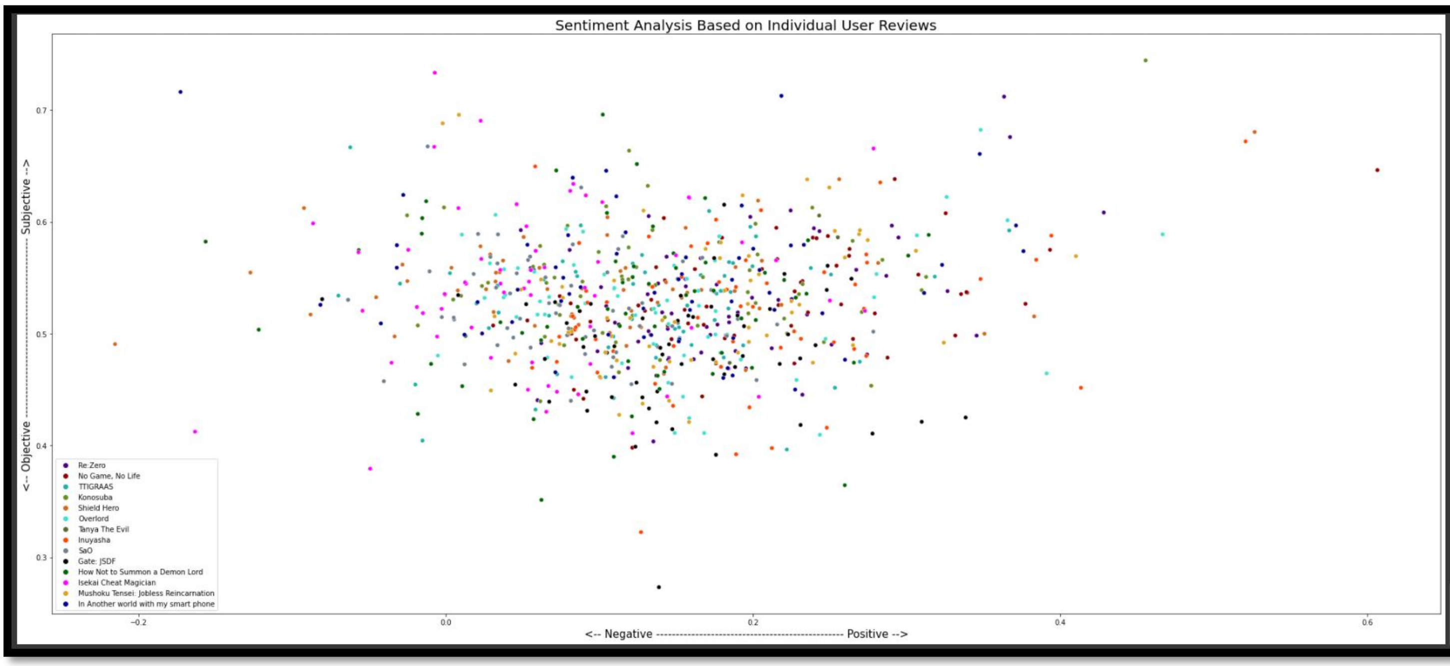
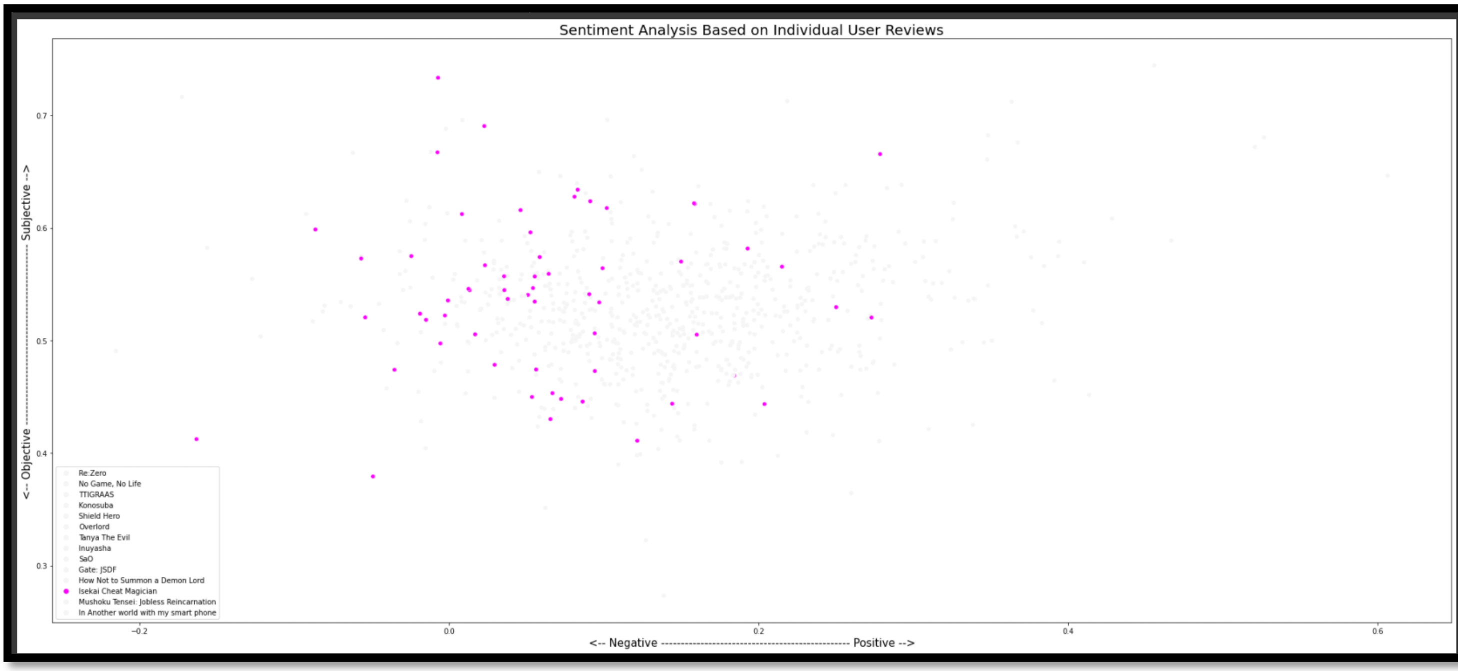


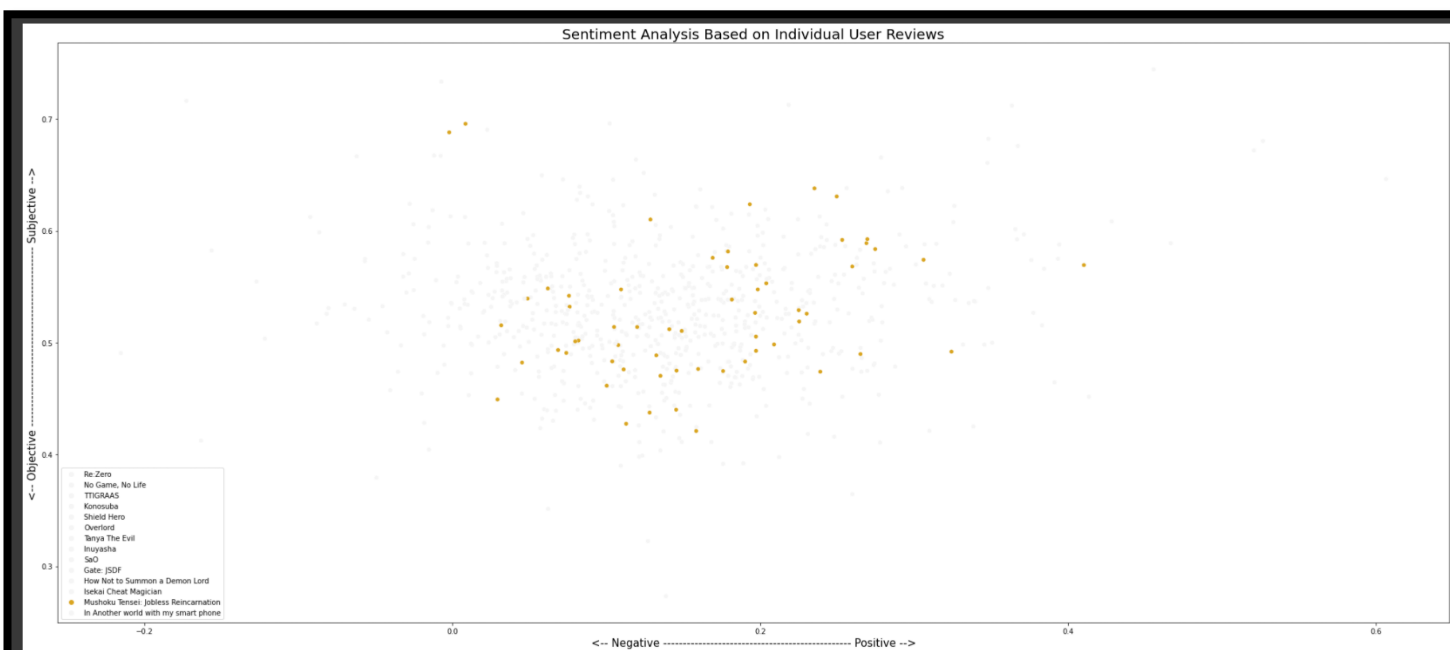
Fig.7 *Legend*

Fig.8. *Sentiment Analysis Plot (Filtered) for Isekai Cheat Magician*



The **Fig.8** plot for (**Isekai Cheat Magician**) shows that subjective opinions are more correlated with negative opinions, as this show reviewed poorly overall. You can observe more negative reviews clustered in the 'Subjective' range, slightly at and a little above 0.5, with a polarity score ranging between 0.0 and 0.1. Although the overall dataset did not show a correlation between negativity and subjectivity. A manual inspection of the reviews for this show reveal that the show is considered 'generic' or 'boring' and that the protagonist is not compelling.

Fig.9. *Sentiment Analysis Plot (Filtered) for Mushoku Tensei: Jobless Reincarnation*



Plotting another show in **Fig.9 (Mushoku Tensei: Jobless Reincarnation)** you can see that most of the clustering happens towards the center of the plot, between 0.0 and 0.2 on the polarity range and slightly above 0.5 for subjectivity. I believe these reviews are relatively middle ground due to how fairly new this anime is, considering its release date was January 2021 and only 11 episodes have aired so far. Thus, in my opinion viewers are reluctant to form strong opinions until they see where the show is going. What is usually seen among fans of a particular series that are book adaptations, if key plot points are missing this will tend to weigh negatively on the show going forward or like in **Fig. 9** where generic and boring stories tend to cause negative reactions. But with this sample data we can see that generally the feedback should often be positive, if we were to inspect some reviews for this series.

Fig.10 *Word Cloud for visualizing most used words.*

In **Fig.10** we create word clouds to show interestingness of our user reviews based on the frequency of words spoken for a given show. We use these word clouds to show what could be possible key concepts that users are discussing / focusing on when giving their review. For example, with **Konosuba**, this show is comedy, but in the word cloud you can also see names such as '**Kazuma**' and '**Aqua**', these are main characters essential to the plot. Most likely the reason these names come up a lot is due to how well their dynamic is in the show given that they tend to bicker quite a bit like brother and sister. *Konosuba* is highly regarded for having well written characters and a fascinating protagonist.

Another show like **SaO (Sword Art Online)** you see can the name of the main character as well: '**Kirito**'. But **SaO** is a show with negative sentiment reviews. We certainly expect a main character's name to show up in a word cloud, but only by combining this with the sentiment analysis can we know if this is due to the character being liked or disliked. The reviews show that Kirito is considered a character who can "do no wrong" and in most cases "can never lose". Viewers complain about this character archetype. Having word clouds gives us another decent heuristic in understanding user sentiment of show which can shed light as to main concepts reviewers focus on be it good or bad.

Fig.11 *Sampled Data*

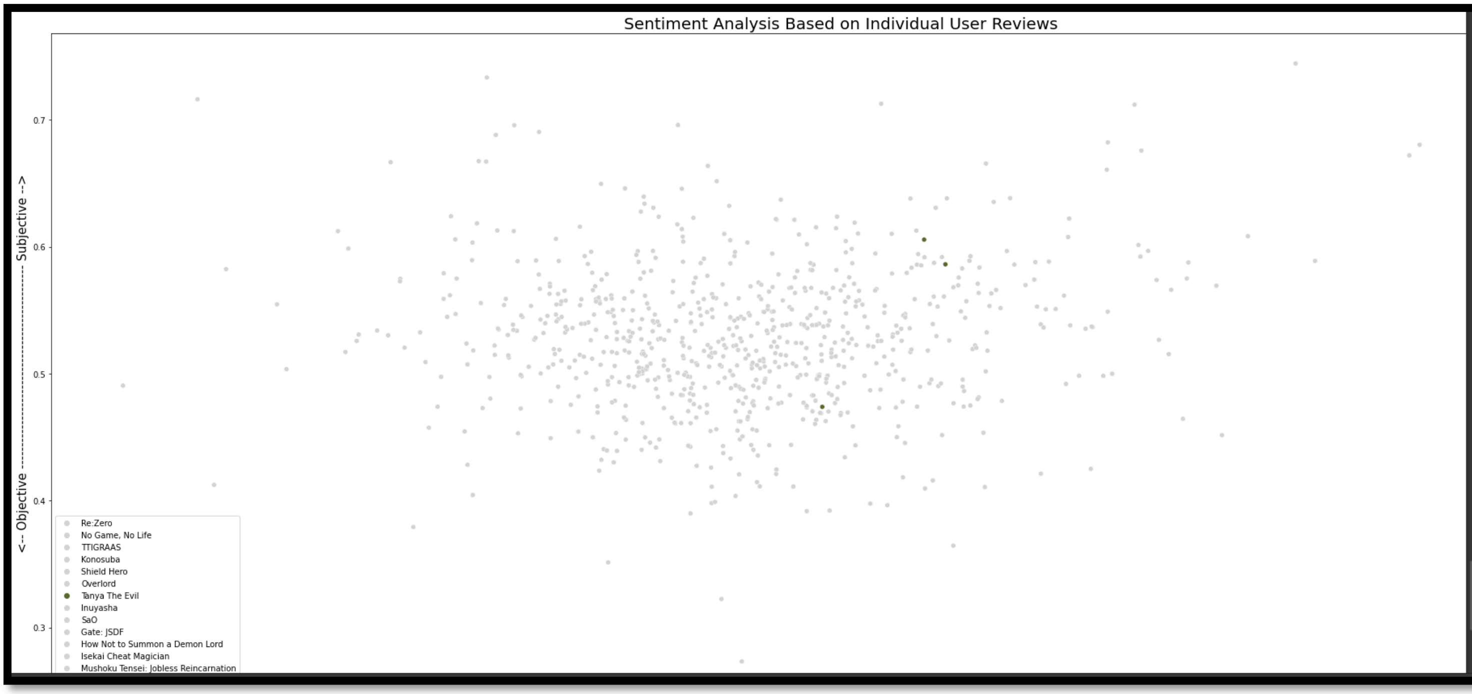
User Review Sample	Show Title	Polarity Score	Subjectivity Score
sorry in advance if i spoil the plot no game no life was a fun anime which i enjoyed even though the core premise is seen way to often in the anime community it had a fun twist on the isekai clich filled with endearing...	No, Game No Life	0.215128	0.594712
you know everytime a new harem show comes out everybodys allways like oh no not again those shows are all the same with boring stereotypes pathetic characters a dreadful story and full of useless fanservice in...	How Not to Summon a Demon Lord	-0.156464	0.582354

Fig.11 *cont.*

User Review Sample	Show Title	Polarity Score	Subjectivity Score
tensei shitara slime datta ken review v i dont get why this anime is loved too much but its nice too see an isekai that follows an optimist and a heartwarming path about comedy the light comedy is liked by the most of	The Time I got Reincarnated as a Slime	0.159929	0.489860
this was an anime that changed my view on the until now mostly boring and unoriginal isekai genre with only a few bighitters and modern classics every now and then with adaptations such as rezero and konosuba	Mushoku Tensei: Jobless Reincarnation	0.158312	0.420946

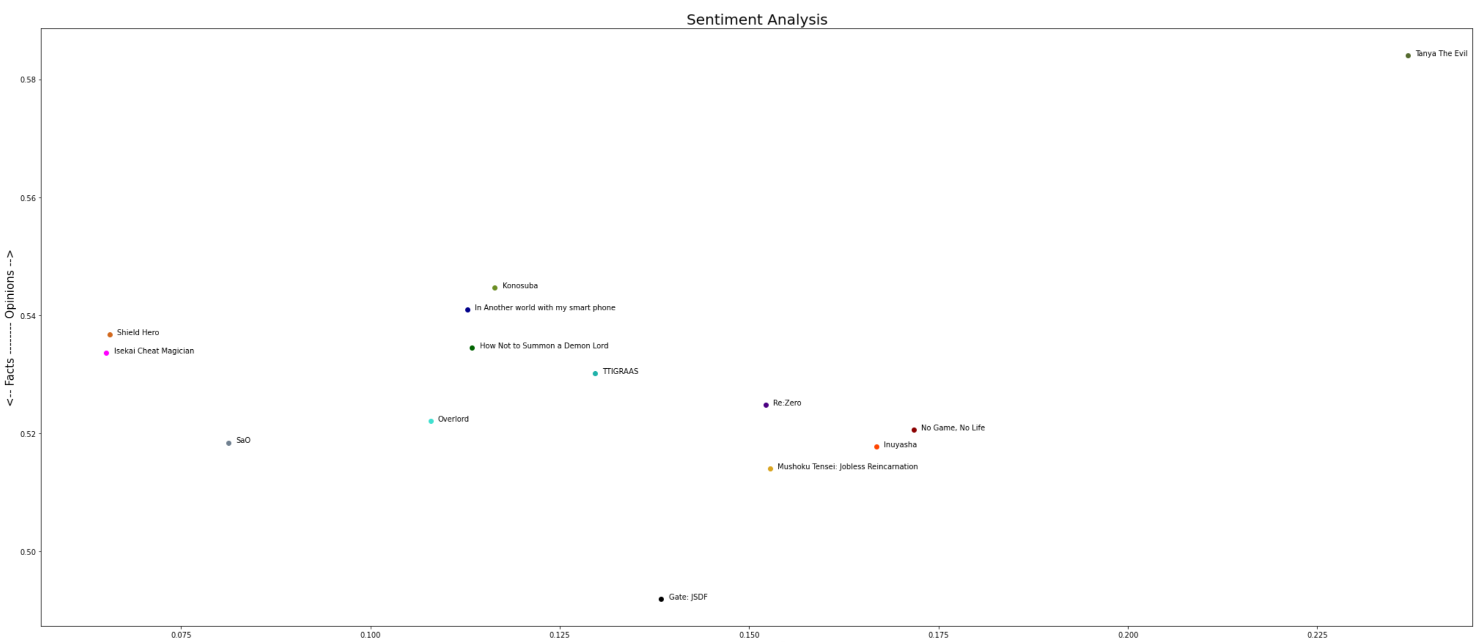
In **Fig.11** you can see some truncated reviews along with the polarity and subjectivity scores. A manual inspection of the review text with the scores shows how the two accord. Some of them start off respectful in the initial approach i.e., the first sample in **Fig.11**. Then you have others which you can clearly determine that this show was not well received by that reviewer i.e., the second sample in **Fig.11**.

Fig.12 *Show* **"Tanya the Evil"** *(green dots)*



This show has positive and subjective reviews, but there are only three reviews recorded that could be found on the site, when compared to **Fig(s). 8-9.**

Fig.13 *Overall sentiment analysis scores for all reviews of each show combined.*



By combing the text of all reviews for a show, we get a sense of the overall sentiment shown towards that show. All shows average out to have a score greater than zero for positivity. Tanya the Evil is an outlier due to only having 3 samples (**Fig.12**).

Conclusion:

To conclude this project, this analysis has given us some good insight as to how user reviews could be leveraged in a way that we can filter through subjective opinions and look more at objective-based opinions. Reviews are plentiful. Passionate show viewers are willing to create reviews that show creators or other viewers can make use of. Of course, a person cannot sift through thousands of reviews to determine how a show is received. Nor does reading a few reviews give confidence in the general reception.

While the reviews did have a numerical review score given by the user, one person's scale will differ from another's. For example, one reviewer may consider anything under 7 bad while another will say that a 5 is average. Giving a show a 6 rating would mean drastically different things for these two reviewers. It is better to rely on what the reviewer actually says about the show, but this is much more difficult to consume quickly.

This project can also show creators and consumers the overall positivity that a show is receiving using a constant scale. This can even be done while the show is still running to gauge if the show is improving or not. We have strong confidence that the sentiment scores generated are accurate. Subjectivity analysis can help find those reviews that give more constructive feedback. Show creators may find such reviews valuable. This program would be able to comb through reviews to find the most objective ones as recommended reading.

Although our pipeline is tailored to the MAL website, it could be refit for any review site to process and plot the overall sentiment for other shows. The limitations for this project come from the site itself. The analysis needs data, the more the better. Some shows did not have many user reviews. Too few reviews can lead to the randomness in viewer opinions coming through too strongly as seen with **Fig.12**.

We limited our data to three pages of reviews, so we could improve accuracy by collecting more reviews. Also, this project could be improved by making the scatter plot interactive to give the user a choice of which show to highlight and a way to view the overall sentiment. Our sentiment analysis is dependent on the pre-tagged values within the TextBlob package. The words are also tagged with part-of-speech data which we do not have for our reviews. Determining part-of-speech for each word might allow the sentiment analysis to be more accurate.

Sources:

[1]: "isekai." *Definitions.net.* STANDS4 LLC, 2021. https://www.definitions.net/definition/isekai.

[2]: Top 20 Isekai Animes (2019 Update) (2019) https://www.myanimeshoponline.com/top-20-isekai-animes/

[3]: Sentiment and Objectivity Classification (Falco X, Witten. R, Zhou, R) (2009) https://nlp.stanford.edu/courses/cs224n/2009/fp/24.pdf

[4]: Source Code for text.en.Sentiments Loria.S (2020) https://textblob.readthedocs.io/en/dev/_modules/textblob/en/sentiments.html

[5]: SUBJECTIVITY LEXICON FOR ENGLISH ADJECTIVES (2014 May 10) https://github.com/clips/pattern/blob/master/pattern/text/en/en-sentiment.xml

[6]: TextBlob Sentiment: Calculating Polarity and Subjectivity (2015 June 7) https://planspace.org/20150607-textblob_sentiment/

[7]: SPSS TUTORIALS: PEARSON CORRELATION
https://libguides.library.kent.edu/SPSS/PearsonCorr