

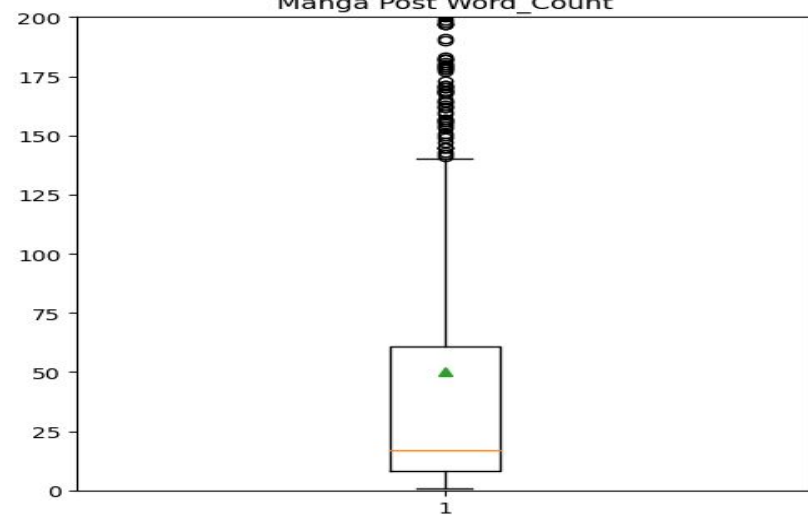


Nihonoku! An NLP Battle between Manga and Anime

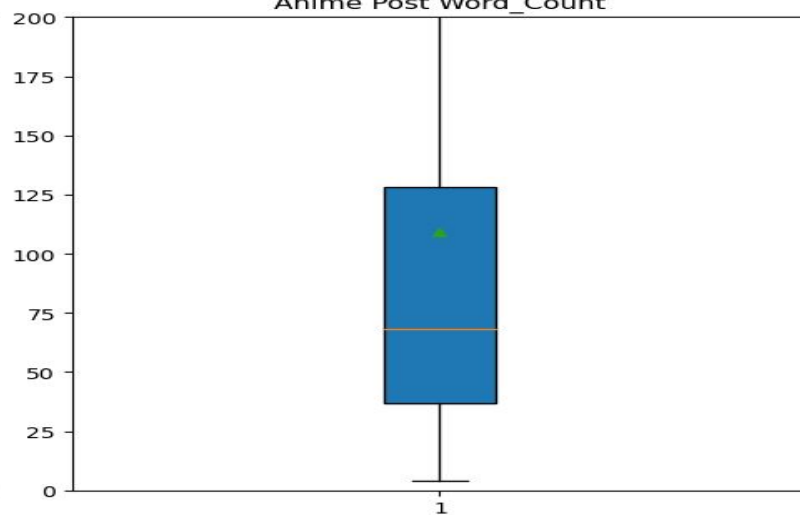
Overview

- Try to build a model to distinguish between the similar mediums of manga (comic books) and anime (TV shows) with the ultimate goal of seeing the relative each manga and anime language can be distinguished to plan for the expected costs to refurbish ads in the future when mangas get adopted into animes.
- The following data was based on over 1600s Reddits from each category, gathered in January 2024.
- Plans originally called for the posts to be the basis of the analysis, being more comparable to ads, but the following roadblock became apparent:

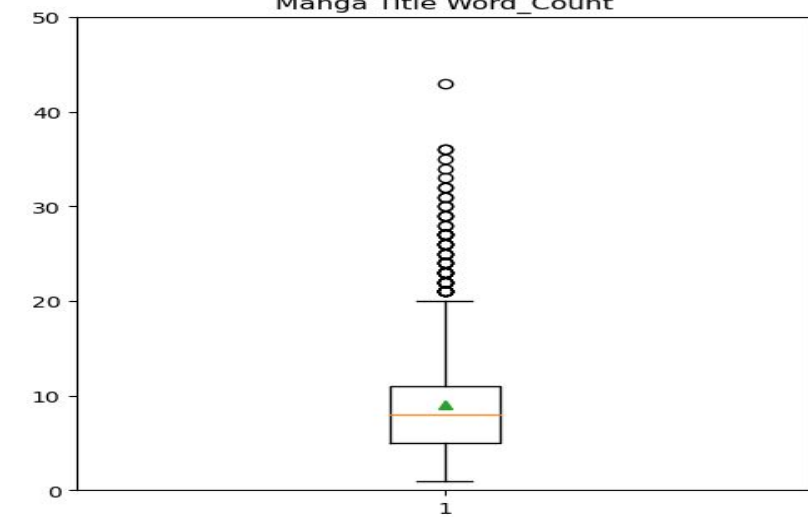
Manga Post Word_Count



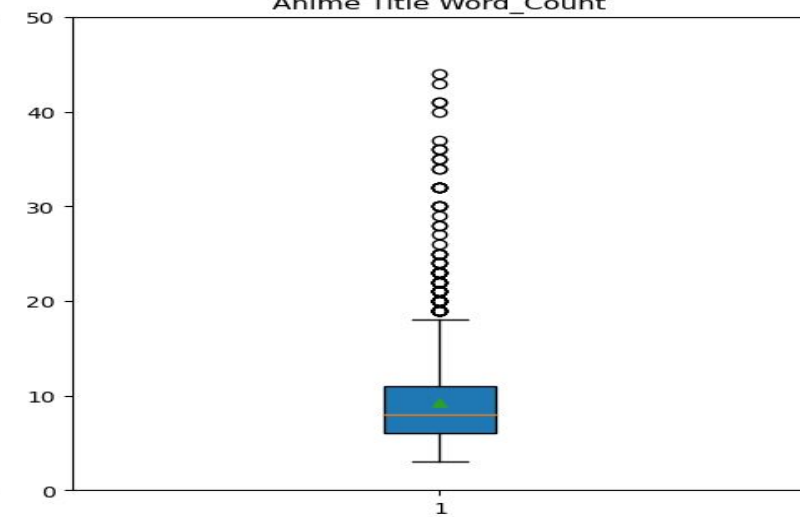
Anime Post Word_Count



Manga Title Word_Count



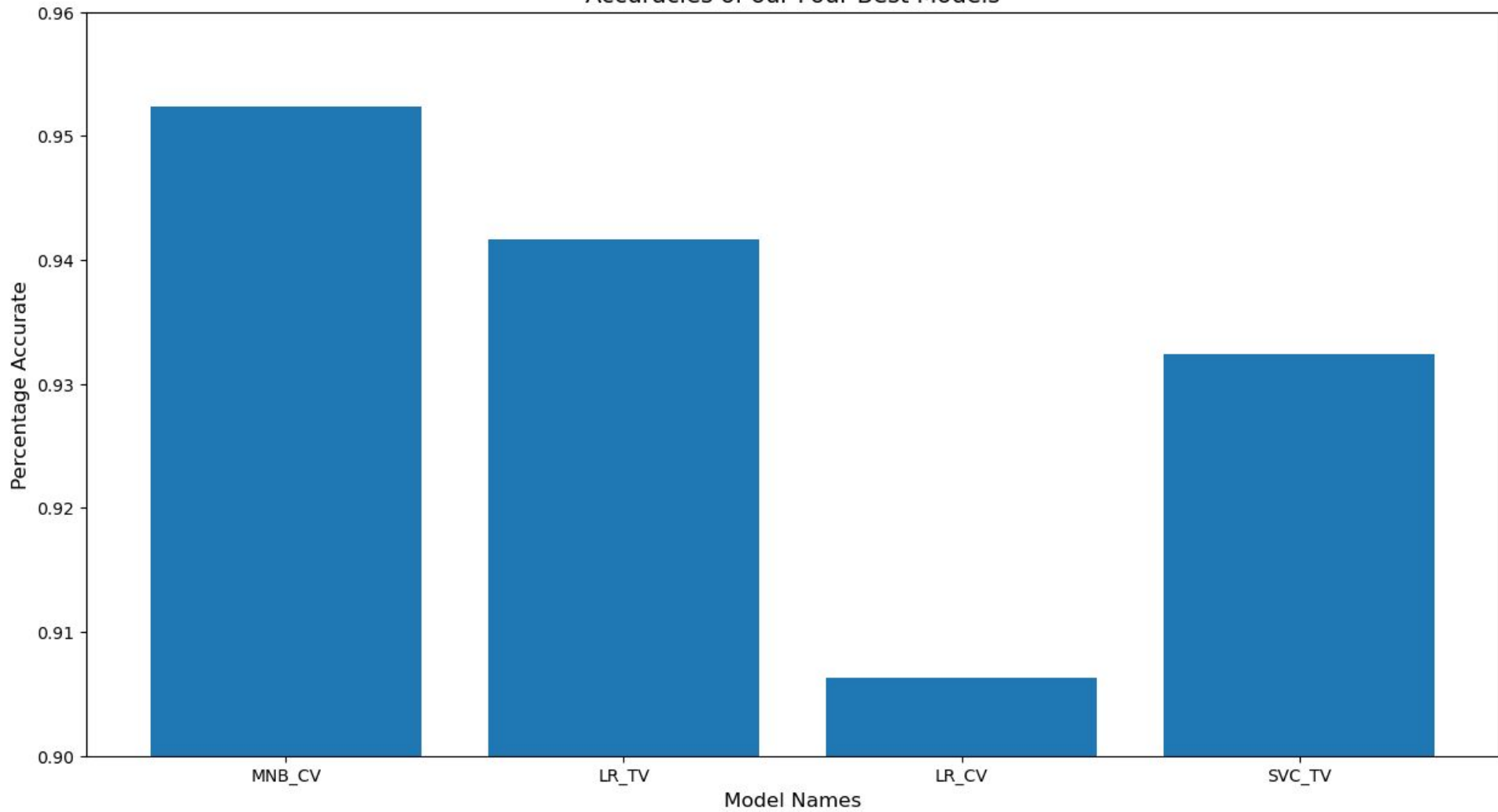
Anime Title Word_Count



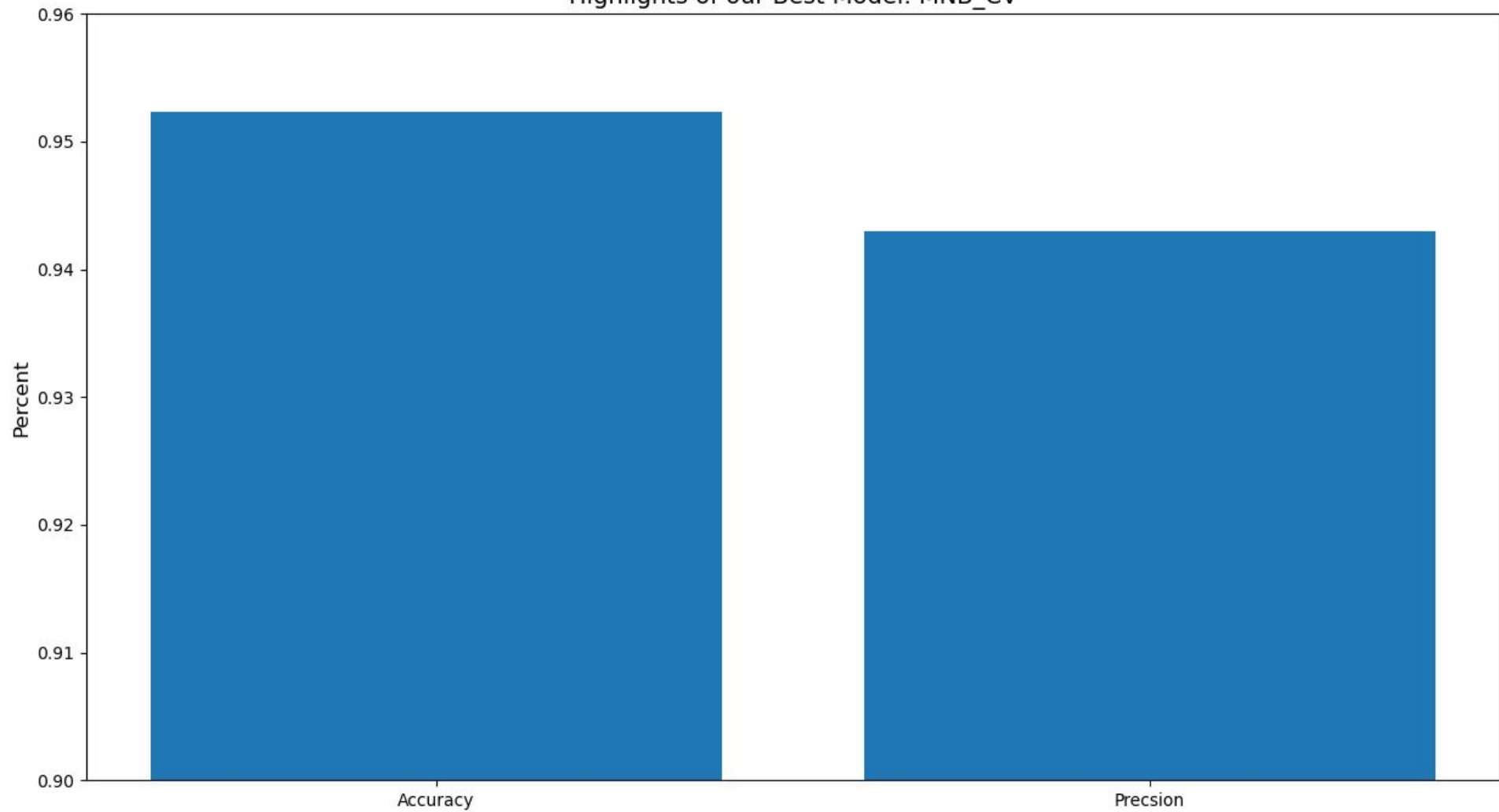
Methodology

- An initial cleaning of non-alphanumeric characters was performed, followed by routine 'simplification' of words, stemming, to try to keep things as simple as possible.
- Accounted for suspected trends in the anime and manga world by being on the lookout for acronyms
- Custom buzz-words, stop words, were explored.
- A total of eight different models were considered, tuning them until they could no longer be made more accurate to classify our data.

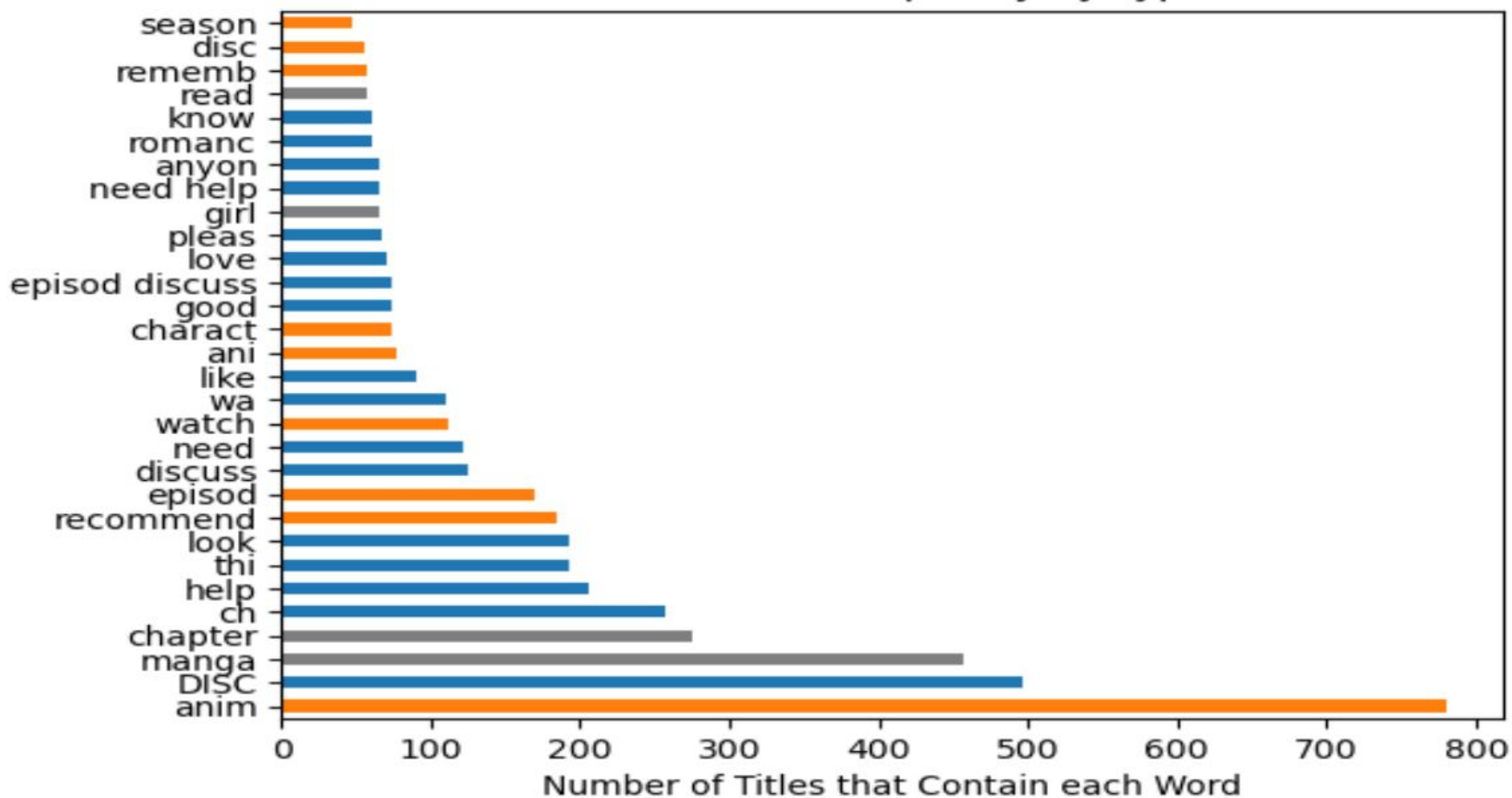
Accuracies of our Four Best Models



Highlights of our Best Model: MNB_CV



Word Count Frequency by Type



A Number of False Positives:

- 'recommend s that are similar to frieren beyond journey s end'
- 'did someth happen to mangaupd'
- 'question look for specif romanc stori'
- 'what newish non WSJ manga without an anim do you think will be a big hit'
- 'can anyon recommend a good adventur manga base on anim I ve watch'

Questions?

