Time Series Model to Predict Future Popular Animes Genres in 2025

Nathanael Setiawan[1*], *Jayson Mikael* Hendra[1], *Bernard* Wijaya[1], *Nunung Nurul* Qomariyah[1], and *Maria Seraphina* Astriani[1]

[1]Computer Science Department, Faculty of Computing and Media, 11480 Bina Nusantara University, Indonesia

Abstract. Abstract Anime is a form of animated media originating from Japan, where it is mostly characterized by its unique style and is exclusively made in Japan. In the current digital era, anime is now very prevalent not only for Japanese consumers, but the entire world. With this much relevance behind anime, this paper proposes some observations of genres which trend over time. This is mostly to help creative teams behind anime studios to appeal to international audiences in the near future using the Prophet time-series. The resulting model has been evaluated with an RMSE of 1.102 and a MAPE of 13.551%. The model predicts that the genres Super Power, Demons, and Supernatural will trend upwards, while the genres Josei, Cars, and Kids will trend downwards into the year 2025.

1 Introduction

Anime is a form of animated media with a distinct anime style alongside usually being produced exclusively in Japan. The community surrounding this form of animation is very large, being extremely popular locally in Japan as well as getting massive attention from Western audiences. Anime has an interesting and long history behind it, with the modern style of anime starting around the 1960s with the advent of Toei Animation [1, 2]. With such a history behind anime, it is unsurprising how deeply rooted Japanese culture is seen in anime, being the country of origin for this unique style as well as being made exclusively for Japanese audiences in the past [3, 4]. However, with such success, as well as being aided by the growth of communication technology, anime is getting extreme amounts of international attention, leading anime studios to reconsider how to appeal to a more global market if they wish to obtain international success.

The appeal of anime will always be subjective, as it is up to the watcher's opinion if a particular anime appeals to them. Because of this, much data around anime is qualitative, which is very difficult to analyze. Fortunately, with tools like MyAnimeList (MAL), we can more easily analyze quantitative data in the form of ratings, with data being provided by MAL, the MAL API, as well as the community around MAL [5]. Alongside providing ratings, MAL also provides an anime's objective information, such as genre, which can be used as independent variables. Another important thing of note is that even though MAL is currently owned by a Japanese owner, most contributing users of MAL are not of Japanese origin, and so the statistics will be of global opinion. For this specific paper, we will attempt to approach our analysis on how the genre attributes affect an anime's popularity based on historical data.

The importance behind this paper is related to both the creators and the consumers of anime. Anime studios need to understand what attributes they need to have to appeal to international audiences going into the future, not only the Japanese demographic [6, 7]. Localization teams also need to predict what recent anime will succeed enough to be worth localizing into their language. Consumers of anime will most likely check the description of anime first before deciding to watch it, so understanding what basic genres are the most attractive to audiences will benefit these groups the most.

This paper uses an approach by using a time-series algorithm to determine future data. The contribution of this paper is a time-series algorithm to utilizes previous data. These data are associated with a date as its independent variable, and forecasts the future data based on the training data. A time-series algorithm is very useful in the case of this paper, as this paper will be observing the trends of these anime genres, which requires a time-series algorithm to accomplish it.

2 Related works

Most research about anime tends to focus on computer vision, or the study of graphical imagery for anime. This is because anime is an art form of sorts, where graphics and visuals are very important. However, because of this, research similar to this paper where time-series are to be utilized is extremely rare, and not much research

* Corresponding author: nathanael.setiawan001@binus.ac.id, jayson.hendra@binus.ac.id, bernard.wijaya002@binus.ac.id, nunung.qomariyah@binus.ac.id, seraphina@binus.ac.id

has been established in this field. A consequence of this rarity has resulted in there not being a proper benchmark to evaluate and compare for this paper. This paper will contribute to that rare field of time-series research as well as provide a starting point for similar research like this in nature. The rarity of this research however does not mean there are no related works.

At the time of writing this paper, authors have identified another paper with a similar premise as our paper. [8] recently published this particular paper. According to their study, they used a similar premise as well as a similar methodology of utilizing reviews from the MAL dataset to train and evaluate a Convolutional Neural Network Model. Although there are similarities, the nature of this paper will be different from the referred paper above. The referred paper seems to focus on creating an accurate Neural Network model, using MAL data to train and evaluate the model, in which the final model has an F1-score of 0.97. While this paper attempts to predict and give direct answers as to what might be popular in the upcoming years by analyzing how the data will trend.

Another research paper that has been published with a less strong connection is titled "Collaborative Recommendation System in Users of Anime Films" [9]. The referenced research paper utilizes the same source for its dataset, Kaggle, to create an anime recommender system. This paper is different from the referenced paper as this paper will create a time-series algorithm and does not recommend anime.

There is another research paper which has created a deep learning recommender system, titled "A Deep Learning Recommender System for Anime" [10]. Again, it utilizes Kaggle for the source of the anime dataset. This referenced paper utilizes CF models, which have been evaluated to be efficient utilizing MSE and MAE. Though it utilizes both the same source of data and the same evaluation technique, this paper utilizes different algorithms, specifically K-nearest neighbors, collaborative filtering, autoencoders, and hybrid models to create the entire recommender system, while this paper utilizes a time-series algorithm in the form of the Prophet algorithm.

Finally, a paper titled "Leveraging Side Information to Anime Recommender System using Deep Learning" has created another recommender system for anime" [11]. It utilizes a collaborative filtering algorithm to recommend users animes. The tool that has been developed in this paper has been shown to have results to be 5% better than SVD Models. This referenced paper is again creating a tool and is similar in premise with the other referenced works; however this paper utilizes a completely different approach by means of time-series algorithms.

3 Methodology

3.1 Data set and processing

This paper would like to acknowledge that the dataset used in this paper has been provided by Marlesson Santana under the CC0 1.0 license and only requested that the source page be cited in the paper. The data itself has been crawled out of MyAnimeList on the 5th of January, 2020, using Marlesson Santana's own crawler program. The file to be used in this paper will be anime.csv, with a size of 16214 unique entries, and includes data in the form of ID, Title, Synopsis, Genre, Aired, Episodes, Members, Popularity, Ranked, Score, Img url, and Link.

Data pre-processing will be detailed next. The first step in doing so is to remove all useless columns from the dataset. This includes but is not limited to, all links and urls, unused IDs, text and synopsis, Profile names, and others which may be removed depending on our needs. The removal is done using Panda's drop function and hard coding the values authors do not require. Authors then remove any rows which contain cells that do not fit the data type, NaN, or NULL values using the dropNA function. Next, the date is formatted from string to datetime, utilizing the datetime function pandas provided. Finally, the genres in the dataset will be split using the split function, to create new entries in the dataset for each genre an anime has, so the tested dataset will have more entries than the raw dataset. The resulting dataset will be formatted in the same manner as Table 1; however none of Table 1's values will reflect the actual values in the dataset, and will be instead just an example.

Table 1. Example of data used.

Index	Aired	Genre	Score
1	Jan, 20 2012	Action	7.23
2	Feb, 29 2003	Action	5.12
3	Dec, 2 2012	Shounen	8.49
4	Feb, 1 2019	Shounen	6.13
5	Mar, 13 2013	Parody	7.95
6	Jun, 30 1990	School	7.32
7	Jan, 29 2001	Comedy	9.95
8	Aug, 22 1970	School	10
9	Jan, 1 1981	Comedy	6.30
10	Oct, 11 1999	Comedy	3.25

3.2 Model and techniques

This study utilizes Prophet, which is a procedure to forecast future data [12]. Prophet itself is an open-source software created by Facebook's data science team and is designed for high amounts of data. Prophet can predict data. It has the other ability, the algorithm can also produce analyses of the data and the predicted data such as analyzing weekly, monthly, and yearly fluctuations, as well as analyzing the trends of the entire dataset including the predicted dataset. For these reasons, the prophet algorithm is perfect to be utilized in this paper.

Before utilizing Prophet, the dataset is separated to create a dictionary, each entry in this dictionary is assigned to a genre as the key, and the value is another dataset which contains all the data of that genre. This is done to format the data according to what is needed for the Prophet algorithm, which requires a dataframe which has only a date and a y value as its two columns.

The user can select a genre to analyze, and after inputting the genre they wish to analyze as well as the

number of years to predict into the future, the Prophet algorithm will then take the dictionary values and predict ahead to the specified years. The prophet algorithm does not need to accept any other parameters other than these to function. This can be repeated by the user multiple times. Authors tested every genre as the first parameter. For years ahead, authors decided to look forward to 2025 as a reference because it can predict a substantial amount of data reliably without overextending too much and use Global Anime Market as the reference (currently the analysis end in 2025).

To evaluate the model, the dictionary will not retrieve the values and data one by one, but rather the dictionary will be iterated over. More dictionaries are applied here to store all the values predicted by the Prophet algorithm. These are stored so that the results can be compared with the actual values, which will be described in the next section of this paper.

3.3 Evaluation method

To evaluate the Prophet model, 2 datasets have been created to train and evaluate the model. Data from before the year 2019 is utilized to train the model and data which ranges from the year 2019-2020 is used to evaluate the model's performance. The training data has 45,479 entries and the evaluation dataset has 1,659 entries, which totals to 47,138.

The performance is measured by multiple evaluation methods for time-series applications. The evaluations are Root Mean Squared Error (RMSE), Normalized RMSE, Mean Squared Error (MSE), Mean Absolute Error (MAE), and finally Mean Absolute Percentage Error (MAPE). The most important values here are RMSE and MAPE, as they are the most commonly used evaluation methods accepted. MAPE can be used to see the average inaccuracy from the forecast provided, while contextualizing the dataset by using a percentage rather than a simple number. Meanwhile, RMSE evaluates the quality of a model by also providing the average error. While it does not have the same property of MSE which penalizes bigger errors due to the exponential operation, it provides more context to the dataset by being on the same units as the data, and RMSE works extremely well with large datasets. The rest of these evaluation methods are to supplement the existing evaluation methods, with variations of RMSE and MAPE that will be used to better understand the performance of the model.

4 Results and discussion

Authors first evaluated the model. The important values to pay attention to are RMSE and MAPE, with other values contextualizing and assisting in the 2 values. The RMSE value is 1.102, with the MAPE value being 13.551%, with a normalized RMSE being 0.186. This can be interpreted as the model being at least 82% accurate. This accuracy is enough to be utilized as a general tool of prediction (>80%); however, it is not enough to create a prediction with pinpoint accuracy (>90%). Fortunately, Prophet has a range of predictions that it outputs, and this paper only utilizes the average between the range. This paper only requires a general overview of predictions.

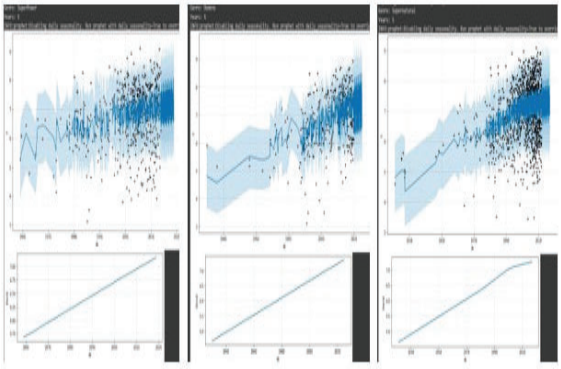

Fig. 1. Left to right: super power, demons, and supernatural results.

This paper will only list notable findings and results, where the trend is significant enough to warrant a place here. While most anime genres no longer trend upwards or their trend is very miniscule, some do trend up a significant amount even towards 2025. The most notable of these are Super Power, Demons, and Supernatural, as seen in Figure 1. On the other hand, a surprising result that has come out of this model has been the revelation that some anime genres seem to have decreased drastically over the past years and have decreased a significant amount. The most notable of these are Josei, Cars, and Kids, as seen in Figure 2.

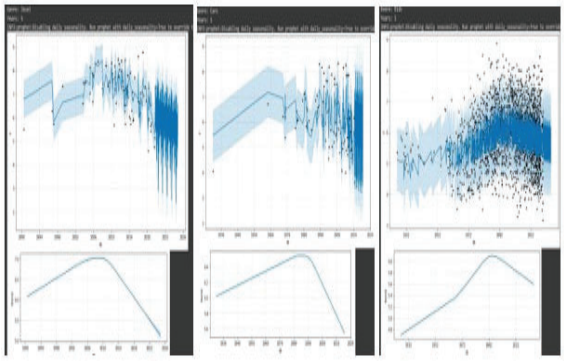

Fig. 2. Left to right: josei, cars, and kids results.

The results have shown interesting trends, especially for the trends which have decreased over the years. First, as stated before, most results show that most genres no longer trend or trend much less upwards or downwards. Most extremely popular anime genres, examples of which are School, Shounen, and Action trends upwards in between the 1970s to the 2000s and 2010s, where they then trend very little if at all around the score of 7 on a 1-10 scale.

The results which have shown an increase in the ratings have a very interesting pattern behind them. The 3 results which are selected for this paper are Super Power, Demons, and Supernatural. They have the common theme of being associated with the supernatural, very similar to those of fantasy. This may

be a result of consumers enjoying more unnatural and more imaginative content rather than the ones that are more grounded in reality. However, this is only a speculation which should be investigated in the future.

Finally, the results which have shown a decrease in ratings also do have an interesting connection. Two of those results, being Josei and Kids, are not only classified as genres, rather they are also demographics. Josei as a genre is supposed to appeal to women who have already graduated from high school. This means that anime made for kids or middle-aged women may not be very popular, especially in the Western market. Though this is another speculation which should also be investigated in the future.

5 Conclusion

In conclusion, the research found that there are some genres which still show growth into the future, most that have stopped growing, and some which are shown to have instead decreased in ratings and are predicted to decrease even more. With an RMSE of 1.102 on a scale of 1-10, and a MAPE value of 13.551%, it can be concluded that a few genres will trend up overtime, the most notable of these are Super Power, Demons, and Supernatural. While there are some that trend downwards, such as Josei, Cars, and Kids. This research has shown that the genres of these anime seem to have a real effect by how they get rated better each year. However, the data has also shown that genre is not the only deciding factor if an anime can get high ratings or not. Ultimately, these animes will be rated badly if they are of low quality, and genre is only one factor in determining the rating of an anime.

Some further research can be done in examining these other factors, such as what do consumers value the most, either animation, story, sound design, etc. Another recommendation is to calculate the statistical significance of the results here, or to utilize another time-series model and compare them with Prophet for either better or worse results.

References

1. Y. Yasuo, *The evolution of the Japanese anime industry* https://www.nippon.com/en/features/h00043/the-evolution-of-the-japanese-anime-industry.html (2022)
2. K. Kirillova, C. Peng, H. Chen, *Anime consumer motivation for anime tourism and how to harness it*, J. Travel & Tourism Marketing **36**, 2, pp. 268-281 (2019)
3. S. Napier, Anime from Akira to princess mononoke: experiencing contemporary Japanese animation (Springer, 2001)
4. J. Berndt, *Anime in academia: representative object, media form, and Japanese studies*, Arts **7**, 4, pp. 56-69 (2018)
5. MyAnimeList.Co.Ltd, https://myanimelist.net/ (2022)
6. L. Carter, *Marketing anime to a global audience: a paratextual analysis of promotional materials from spirited away*, East Asian J. Popular Culture **4**, 1, pp. 47-59 (2018)
7. R. Mihara, *A coming of age in the anthropological study of anime?* J. Business Anthropology **9**, 1, pp. 88-110 (2020)
8. S. M. AlSulaim, A. M. Qamar, *Prediction of anime series' success using sentiment analysis and deep learning*, 2021 International Conference of Women in Data Science at Taif University (WiDSTaif), pp. 1–6 (2021)
9. A. S. Girsang, B. Al Faruq, H. R. Herlianto, S. Simbolon, *Collaborative recommendation system in users of anime films*, J. Physics: Conference Series **1566**, 1, pp. 012057 (2020)
10. V. M. Mutteppagol, A deep learning recommender system for anime, Ph.D. dissertation (National College of Ireland, Dublin, 2021)
11. A. T. Wibowo, *Leveraging side information to anime recommender system using deep learning*, 2020 3rd International Seminar on Research of Information Technology and Intelligent Systems (ISRITI), pp. 62-67 (2020)
12. J. Taylor, B. Letham, Forecasting at scale (PeerJPreprints, 2017)