International Conference on Recent Trends in Data Science and its Applications
DOI: rp-9788770040723.013

Analysing Twitch Streamer's Success from Twitter

A. ParveenAkhther
Department of Computer Science & Engineering Sathyabama Institute of Science & Technology,
Chennai – 600 119,Indiaparveena77@gmail.com

A.Maryposonia
Department of Computer Science & EngineeringSathyabama Institute of Science & Technology,
Chennai – 600 119, India
maryposonia.cse@sathyabama.ac.in

Daniel Lawrence
Department of Computer Science & Engineering Sathyabama Institute of Science & Technology,
Chennai – 600 119, India
daniellawrence4150@gmail.com

Arun Kumar Kailash G
Department of Computer Science & Engineering Sathyabama Institute of Science & Technology,
Chennai – 600 119, India
arunkkailash@gmail.com

Abstract— Twitch is a platform that provides live streaming services for videogames; it has created a great environment for people that enjoys both watching and playing games, but with the increase in the number of streamers and games, streamers must find ways to get better recognition and find all points of leverage for better interaction on their streams and content. This paper analyzes the relationship between a Twitch streamer's success with their presence on Twitter

Keywords — social media, social media integration, prediction, random forest; decision tree, Twitch, Twitter, data science, machine learning

I. Introduction

Twitter provides free microblogging and social networking service. Twitter is one of the top 5 social media applications that is used worldwide. Twitter users can broadcast short posts which are known as "tweets", The tweets can be in the form of text, videos, photos, and links. People can get the latest updates and promotions from brands. More than 100 million users post 340 million tweets daily.Since the start of Twitter, academics and businesses have tried to find ways to gain intel and insights from it.

Twitch is a platform that provides live streaming services for video game streaming and other interactions. This paper deals with showing the relation between twitch streamers' presence on Twitter and the success of twitch streamers.

II. Data

A. *Dataset*

The data used in this analysis was taken from Kaggle, "Top 8800 Twitch Streamers" [1]: - The dataset contains information about the top 8800 twitch streamers that are currently streaming on the platform.

This dataset covers some essential criteria that are used to gauge the performance of a streamer.

B. *Exploring the Dataset*

The dataset about the streamers contains the following features as given in Figure 1.

attribute	description
top count	the ranking of the streamer based on watch time (disscussed later)
screen name	name of the streamer
watch time	total number of minutes people have consumed his content
stream time	number of hours he has streamed
peak viewers	number of peak viewers
average viewers	number of viewers the streamer receives on average
followers	number of followers the streamer currently has following his content
followers gained	number of followers gained in the last month
views gained	number of new viewers gained in the last month
partnered	whether or not the streamer is partnered with twitch
mature	whether or not the streamer's content is mature in nature

II. DATA

Fig. 1. All Features and their description

Since the dataset contains a few extra features which are not needed in this project a few features were taken and a few removed as given in Figure 2.

DOI: rp-9788770040723.013

attribute		reason
top count	removed	since it is not relevant to the analysis
screen name	kept	for sentiment analysis on twitter
watch time	kept	it is a very important feature
stream time	kept	since it is only variable the streamer has a control over
peak viewers	removed	since it is a one-time number and does not affect follower and viewers
average viewers	kept	the most important feature,as it leads to monetization and income for the streamer
followers	kept	to see its impact on average views
followers gained	removed	since it is not relevant to the analysis
views gained	removed	since it is not relevant to the analysis
partnered	removed	since it is not relevant to the analysis
mature	removed	since it is not relevant to the analysis

Fig. 2. All Features and reasons for keeping them

The attribute that decides which streamer is at the highest ranking is the watch time as it generates the mostrevenue and is given more attention by the algorithm,governing a streamer's appearance on a viewer's watch feed.

A variety of aspects affect a streamer's watch time, but there is one which is considered crucial among social media personalities and that is their social media presence, particularly Twitter. This is considered important as it effects how many people they are able to attract to their stream to view their content.

This presence is not direct, that is someone with a positive presence is not necessarily going to have the highest watch time and similarly people with the most negative presence may not have the lowest watch time.

So, in this paper we analyzed this aspect by using data from Twitter using Tweepy [11], the Twitter API. It filters Tweets made about a specific streamer in the past seven days and segregates them into positive, negative or neutral in nature.

III. Related Work

Many researchers have tried to find relations between the entertainment industry and social media, and how their presence on social media affects either their job or their popularity.

In [2] Dr.Rose Catherine aims to accurately predict the box office success of upcoming movies using tweets.

In [3] J. Huang, W. F. Boh, and K. H. Goh investigate the relationships between comments generated from social media and sales using a natural language processor, similarly in [4] and [8] by using meta-data mined from social media such as number of likes on a Facebook page for the movie, follower count of actorson Twitter and the number of likes on the trailer of themovie on YouTube.

In [9] Vasu Jain's analysis tweets about movies to predict several aspects of the movie popularity and covers how movies have a higher rate of success based on the visibility of the movie on social media help in the sales of the movie.

When twitch, a primarily game streaming platform is being considered it is understandable to consider games and their trends. The present technological boom has to lead to a major increase in people that enjoy both playing and watching video games.

In [5] Boris Bankov has written about the impact of social media on video game communities and the gaming industry. It describes the evolution of the gaming industry and the impact of integrating social media subsystems on gaming communities. The study includes a relationship between social media platforms and the gamingindustry.

[12] is a very highly recommended article about the author, Henry Jenkins teaching a course on Transmedia Entertainment and Storytelling, similarly [13] is also an article by the same author discussing the impact social media has on businesses.

[6] talks about trying to make predictions from socialmedia, and points out the many pitfalls, also the past methods used to do so.

In [10] ThienHai Nguyen and KiyoakiShirai build a model to predict stock price movement using sentiments on social media and subsequently describethe relation between the two. Both [14] and[15] analyze Twitter for predicting results of major events with a large crowds, that being Elections and the results of a football match in a league respectively.

IV. Approach

In a broad sense, the approach is to perform sentiment analysis on Twitter data corresponding to the respective twitch streamer and analyze how it impacts the streamers' watch time since it is deemed the most important figure to gauge a streamer's success. The first step is to clean the data. The dataset contains a lot of streamer names in other languages and not in English, hence it is not easy to process. Those names need to be replaced with their respective Twitterhandles.

By performing the sentiment analysis we get the following new columns as given in Figure 3.

attribute	description
positive	number of positive comment about the streamer
negative	number of negative comment about the streamer
neutral	number of neutral comment about the streamer

Fig 3:.New columns to the data

DOI: rp-9788770040723.013

Before segregating the results of the sentiment analysis to the respective column the columns look as follows:

s no	screen name	stream time	followers	watch time	positive	negative	neutral	average viewers
0	xQcOW	215670	3691010	7333609065	0	0	0	32913
1	Gaules	515595	1966465	6314532585	0	0	0	12254
2	summit1g	216000	5374710	6235007490	0	0	0	25931
3	ESL_CSGO	517965	4195657	4764929775	0	0	0	9249
4	NICKMERCS	131880	4415637	3853252845	0	0	0	29183

Fig 4: Data Frame before sentiment analysis is performed

The data in the new column will be zero as given in Figure 4, sentiment analysis has not been performed, hence these columns are left empty. Sentiment analysis is performed on all the streamer tags and increments the respective sentiment category. In [7] and [8] the type of sentiment analysis techniques is covered.

s no	stream time	followers	watch time	positive	negative	neutral	average viewers
0	215670	3691010	7333609065	8	2	20	32913
1	515595	1966465	6314532585	8	12	80	12254
2	216000	5374710	6235007490	10	22	66	25931
3	517965	4195657	4764929775	26	4	11	9249
4	131880	4415637	3853252845	51	9	40	29183

Fig 5: Data Frame after sentiment analysis is performed

Once sentiment analysis is done there is no need for the streamer tags, hence they can be removed.

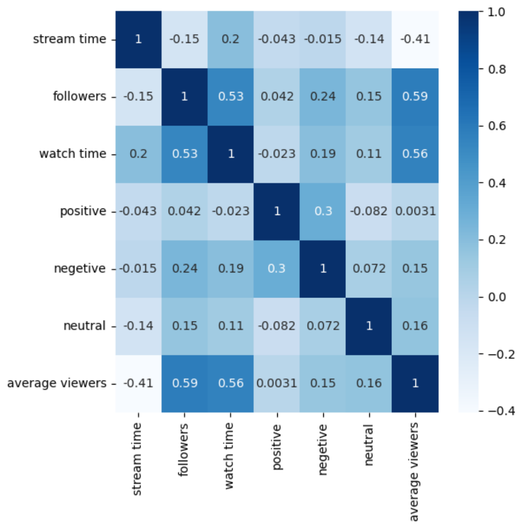

Fig 6: correlation matrix

Since watch time is the most important feature to focus on as it is the feature that determines the success of the streamer, we train a random forest regressor on it, since Random Forest Regression is a powerful and accurate machine learning algorithm that is used for both regressionand classification type problems.

From figure 6 we can understand that followers and Averageviewers are more related to Watch Time out of the three important features ('stream time', 'followers', and 'average viewers'). Hence, we train the model on the data described in fig 5 leaving out the stream time feature and leaving out the:

i)positive and neutral features ii)negative and neutral features iii)positive and negative features

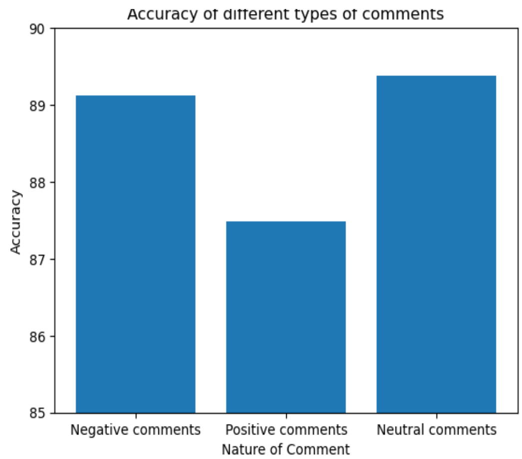

Fig. 7. Accuracy of different types of comments models

Figure 7 gives the accuracy of the random forest regression model trained on the previously mentioned criteria. The model's accuracy is at a minimum, over 87% for positive type comments, and a maximum, over 89% for negative andneutral type.

IV. Conclusion

In this work, we presented an analysis of the relationship between the presence of a Twitch streamer on Twitter and the streamer's success, based on how the streamer is perceived on Twitter. We know that a streamer is considered successful based on the streamer's watch time. We found that the Watch Time is most related to Followers and Average Viewers.We also foundbased on the performance of our model that negative and neutral comments have more impact on the Watch Time.

References

[1] Kaggledataset:- Top8800TwitchStreamers,"https://www.kaggle.com/datasets/girlazo/top-8800-twitch-streamers"

[2] Rose Catherine and SnehaChaudhari, "Predicting Movie Success from Tweets"

[3] J. Huang, W. F. Boh, and K. H. Goh, "From a social influence perspective: The impact of social media on movie sales"

[4] Shruti, S. Roy, and W. Zeng, "Influence of social media on performance of movies".

[5] Boris Bankov, "The Impact of Social Media on Video Game Communities and the Gaming Industry".

[6] Harald Schoen, Daniel Gayo, PanagiotisTakis Metaxas, EniMustafaraj, Markus Strohmaier, and Peter Gloor, "The Power of Prediction with Social Media"

[7] Sitharthan, R., Vimal, S., Verma, A., Karthikeyan, M., Dhanabalan, S. S., Prabaharan, N., ...&Eswaran, T. (2023). Smart microgrid with the internet of things for adequate energy management and analysis. Computers and Electrical Engineering, 106, 108556.

[8] AdityaBaravkar, RishabhJaiswal and JayeshChhoriya, "Sentimental Analysis of YouTube Videos.

[9] Vasu Jain, "Prediction of Movie Success using Sentiment Analysis of Tweets".

[10] ThienHai Nguyen, KiyoakiShirai, "Topic Modeling based Sentiment Analysis on Social Media for Stock Market Prediction"

[11] Tweepy library. http://tweepy.github.com/

DOI: rp-9788770040723.013

[12] Henry Jenkins, "Transmedia Storytelling", Technology Review, p. 1, January 2003.

[13] Henry Jenkins, "Welcome to Convergence Culture", Receiver, p. 2, March 2005.

[14] Moshika, A., Thirumaran, M., Natarajan, B., Andal, K., Sambasivam, G., &Manoharan, R. (2021). Vulnerability assessment in heterogeneous web environment using probabilistic arithmetic automata. IEEE Access, 9, 74659-74673..

[15] H. Rue, and O. Salvesen, "Prediction and retrospective analysis of soccer matches in a league"