

WeRateDogs Twitter Data Analysis

Beshir Aman

Term2: Project 3

The aim of this report is to analyze cleaned WeRateDogs Twitter data. WeRateDogs is a twitter account that comment on dogs in a humorous way and give them ratings where denominator is almost always 10. What is interesting, the numerator is quite often greater than or equal to 10 which indicates how people admire these dogs. After wrangling data, the final dataset includes 1971 unique tweets and 17 features collected from Nov 15, 2015 to Aug 1, 2017.

In Figure 1, we see retweet count over time on the left-hand side and favorite count time series on the right-hand side. In general, the account is gaining popularity with time. The favorites was increasing quicker than retweets. There are some tweets where it was insanely popular. According to Table 1, the tweet that got the highest number of favorites (143,931 likes) was “[Here's a super supportive puppo participating in the Toronto #WomensMarch today. 13/10](#)”. The tweet that got retweeted the most (77,712 retweets) was “[Here's a doggo realizing you can stand in a pool. 13/10 enlightened af \(vid by Tina Conrad\)](#)”.

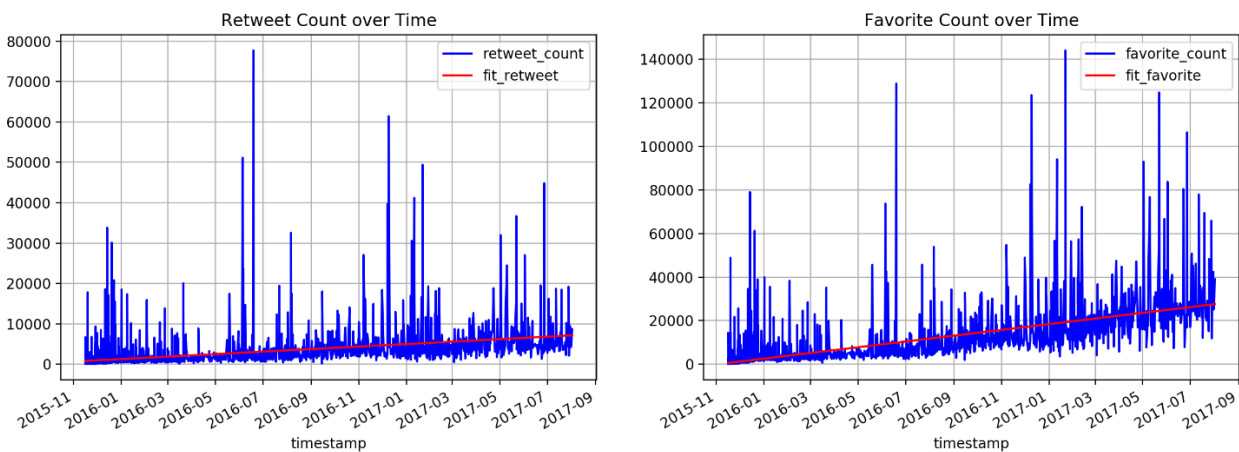


Figure 1

	retweet_count	favorite_count
count	1971.000000	1971.000000
mean	2749.910705	8936.610350
std	4738.288178	12665.843933
min	13.000000	80.000000
25%	613.500000	1953.500000
50%	1333.000000	4067.000000
75%	3156.000000	11241.000000
max	77712.000000	143931.000000

Table 1

Figure 2 illustrates a high linear correlation between retweet and favorite counts. The more you have retweets, the higher likes you get.

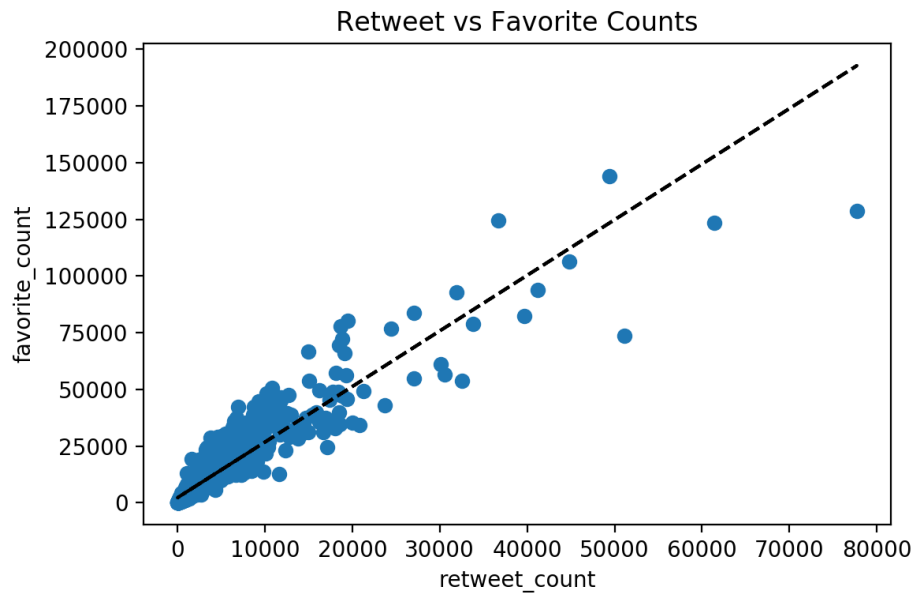


Figure 2

Each tweet is associated with ratings that have denominators and numerators. The denominators are mostly 10, but there are few with different values. The numerators can be any value but mostly they are between 10 and 13. Figure 3 highlights rating ratio over time. Few tweets were dramatically high in rate. At early time, the ratings were generally low (less than one). As time passes, the ratings were apparently increasing. This can be due to increase of account popularity.

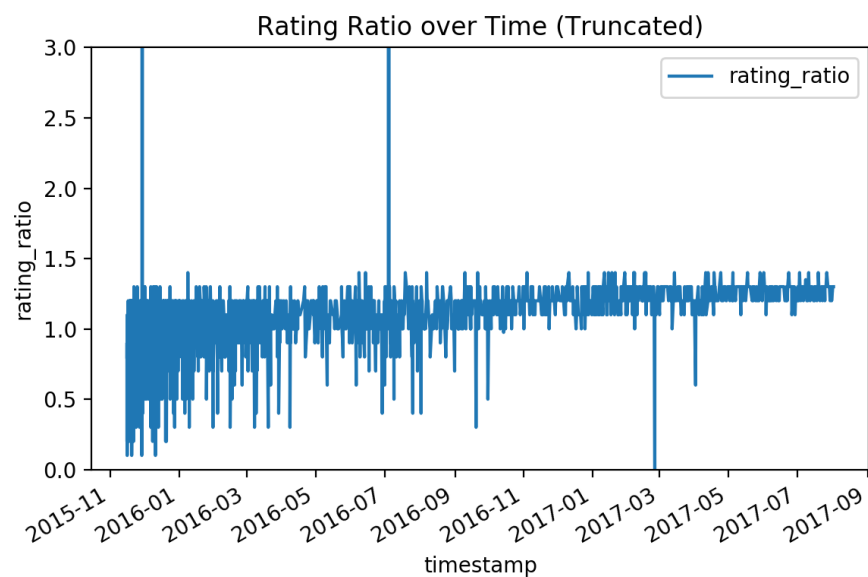


Figure 3

Dog stage information was mentioned in about 300 tweet's texts, including: Pupper, Doggo, Puppo and Floofer. Figure 4, pie chart show that Pupper dogs are the most famous with about 67% appearance, Doggo came second by 21%. Four tweets had both Doggo and Pupper dogs.

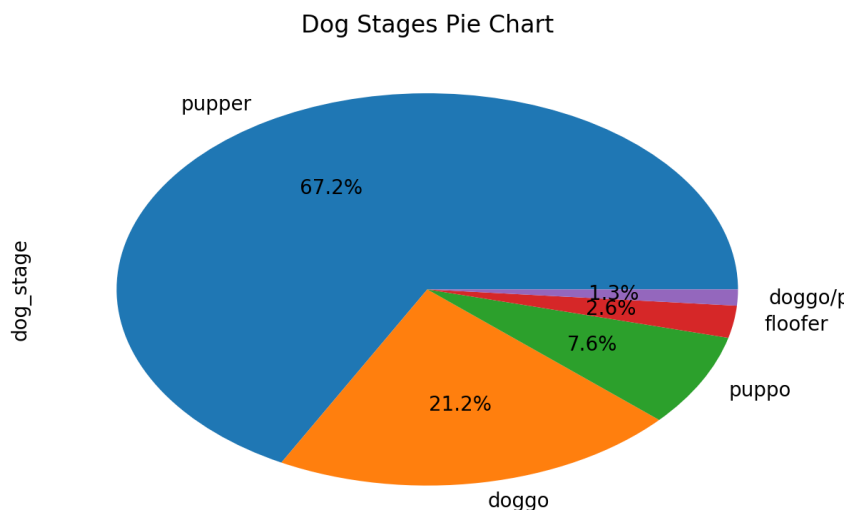


Figure 4

A neural network prediction algorithm results was provided for WeRateDogs tweets. Results include top three predictions whether an image is dog or not. If yes, the dog breed was presented along with confidence level. Figure 5, demonstrates the top dog breed prediction among three choices with corresponding confidence level for each tweet. 1666 out of 1971 was successfully classified to a dog breed, and the rest were not properly predicted. In below figure, top 10 dog breeds prediction are shown. Golden Retriever and Labrador Retriever are the most common breeds.

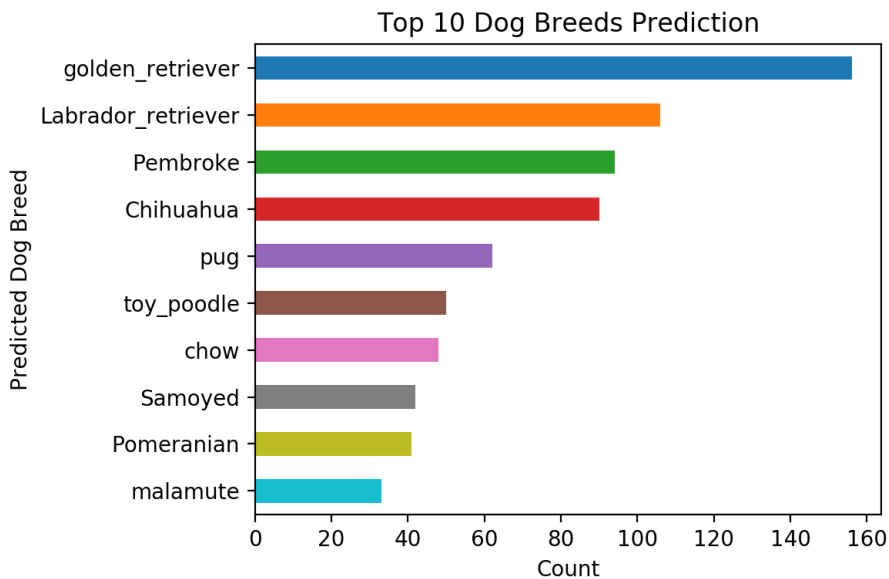


Figure 5