

act_report

December 16, 2018

0.0.1 Data Insights and Conclusions

Tweet and associated data from three sources were collected, cleaned and analyzed to obtain meaningful insights from the twitter archive of the user WeRateDogs. After merging the three dataframes representing the three sources, and fixing the cleanliness issues, it was seen that there were 2356 instances and 19 attributes. These were the tweet_id, timestamp, source, text, rating_numerator, rating_denominator, name, dog_type, favorite_count, retweet_count and the 9 image prediction details columns.

Of the 4 sources used to render the images, iPhone, Vine, Web_client and Tweetdeck, iPhone was found to be the most frequent, with 2221 source listings.

There were 690, 1183 and 483 tweets for the years 2015, 2016 and 2017 respectively. The maximum number of tweets during 2016 is expected since the tweets were collected from November 2015 to August 2017.

From the information contained in the texts, the dogs were classified into dog stages, ranging from pupper to doggo to floofer, and alternative intermediate or composite stages. The pupper was seen to be the most frequent stage among the classifiable dogs (from the tweet texts) of the categories encountered viz. 'pupper', 'doggo', 'puppo', 'doggopupper', 'floofer', 'doggopuppo' and 'doggofloofer'.

The highest favorite count was 140764, for an unnamed puppo, apparently a super supportive one participating in the Women's march in Toronto. The highest retweet count was 61150, for Stephan, a helpful dog predicted to be a Chihuahua with 50% confidence level. A scatter plot analysis showed that there was a strong correlation between the favorite and retweet counts for the dogs.

The dogs received ratings between 11 and 15. The most frequent ratings were in the range of 11-12, numbering around 900.

The average confidence levels for dog predictions based on their images were 60% for their first prediction, and lower for the subsequent two breed predictions.