

Analysis and Insights on WeRateDog

Based on AmnicoreAgency's statistics of Twitter, the number of tweets sent per day is 500 M! It certainly indicates the usage and the lifestyle of the population nowadays. The project aims to emphasize how to utilize the amounts of tweet data through Twitter API. It features an access to exploit the Twitter data of the account @dog_rates, aka WeRateDogs. WeRateDogs has over 4 million followers and has it's popular for dog lovers community. It has received international media coverage. The idea of WeRateDogs profile is to rate people's dogs with a good-natured comment about the dog. The rating system is based on a fraction, with the denominator fixed at 10 and the numerator is almost always a number greater than 10. You must question the judging criteria but because "they're good dogs Brent."

For this analysis I collected data from three different sources. WeRateDogs gave Udacity exclusive access to their Twitter archive for this project in the form of a csv file. This archive contains basic tweet information (tweet ID, timestamp, text, etc.) for all 5000+ of their tweets. Each tweet image was run through a convolutional neural network aiming to analyze the images To correctly identify the dog breeds. The CNN predictions were programmatically downloaded using the Requests Python library as a tsv file. Lastly, using the tweet IDs from the WeRateDogs archive I run queries to the Twitter API to get tweet's JSON data using the Python's Tweepy library. The JSON data of each tweet are kept in dictionary, which I would later use to analyze the tweet's retweet and favorite counts. The analysis part had exploratory questions of the WeRateDogs data. Like what are the most 5 common dog names in the dataset? Top 5 dog names are Oliver, Cooper and charlie ten times each, which dog has the lowest rating? .. etc

In Summary, the account holder of WeRateDogs prefers tweeting from an iPhone. Also the most common dog is Oliver and the highest occurrence of dog doggolingo (type) is pupper and its value is 203. The relationship between favorites and retweets is positively correlated.

Meaad AlRshoud

Feb 2019

