We Rate Dogs

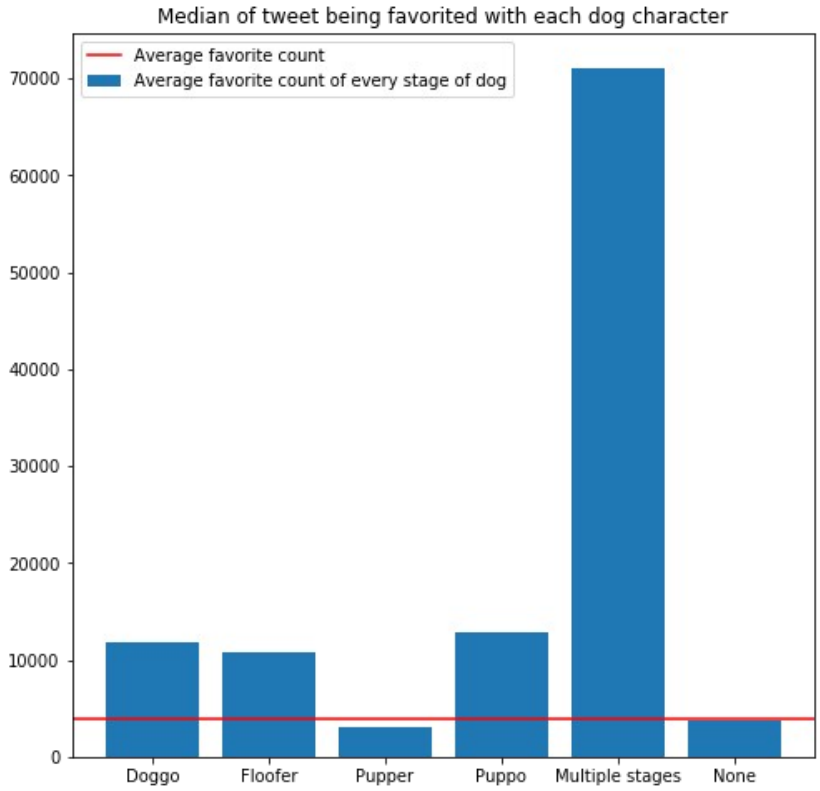
This data set contains data of 1660 WeRateDogs tweets since its beginning (November 2015) to August 2018. WeRateDogs is a Twitter account that rates people's dogs with a humorous comment about the dog and score usually above 10/10. It was started in 2015 by college student Matt Nelson, and has received international media coverage for its popularity. It has acquired more than 7 million followers since its debut.

This data has 27 columns and it can be sliced in to 3 main categories of columns:

- *General tweet data*. Columns in this category are unique tweet id, date, time, full_text, hastags and links used in the tweets, source, retweet and favorite counts, etc.;
- *Unique WeRateDogs data*. This data set contains rating values as well as given dog associasions (doggo", "floofer", "puppo" or "pupper");
- *Computer Neural network algorithm data.* Computer algorithm tries to predict dog breed from the pictures of tweets. This data also contains indicators which shows if algorithm value is true or not and how 'confident' an algorithm is about its decision.

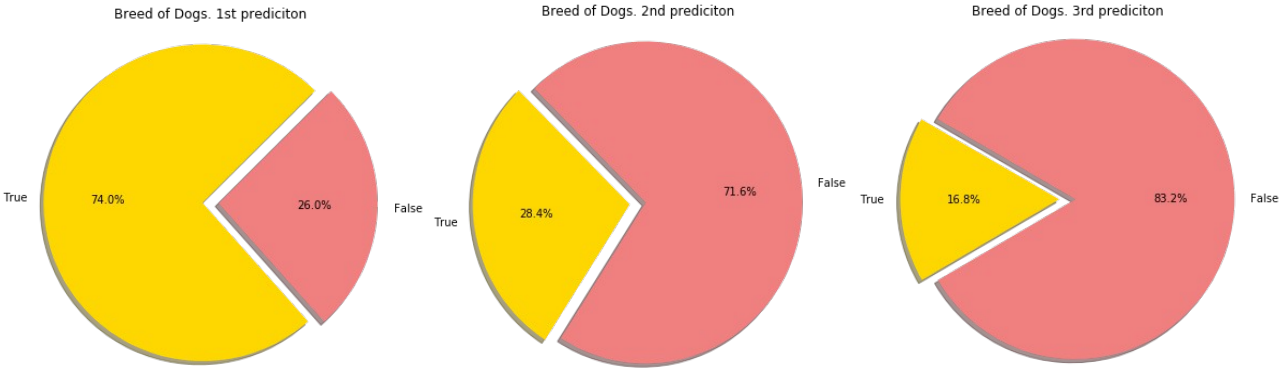
Using small portion of explanatory data analysis we can some interesting findings in this data. For example if we look to dog character associations and compare amount of tweet being favorited or just compare a performance of neural network we can discover these observations:

- Tweets which dog associated with "doggo", "floofer", "puppo" and multiple stages characters tend to have higher favorited counts. However, "pupper" seems to have lower number of being favorited. More investigation has to be done determine, is these dog "characters"actually influences favorites. Also there are low number of tweets with dog_stages compared to tweets without them.



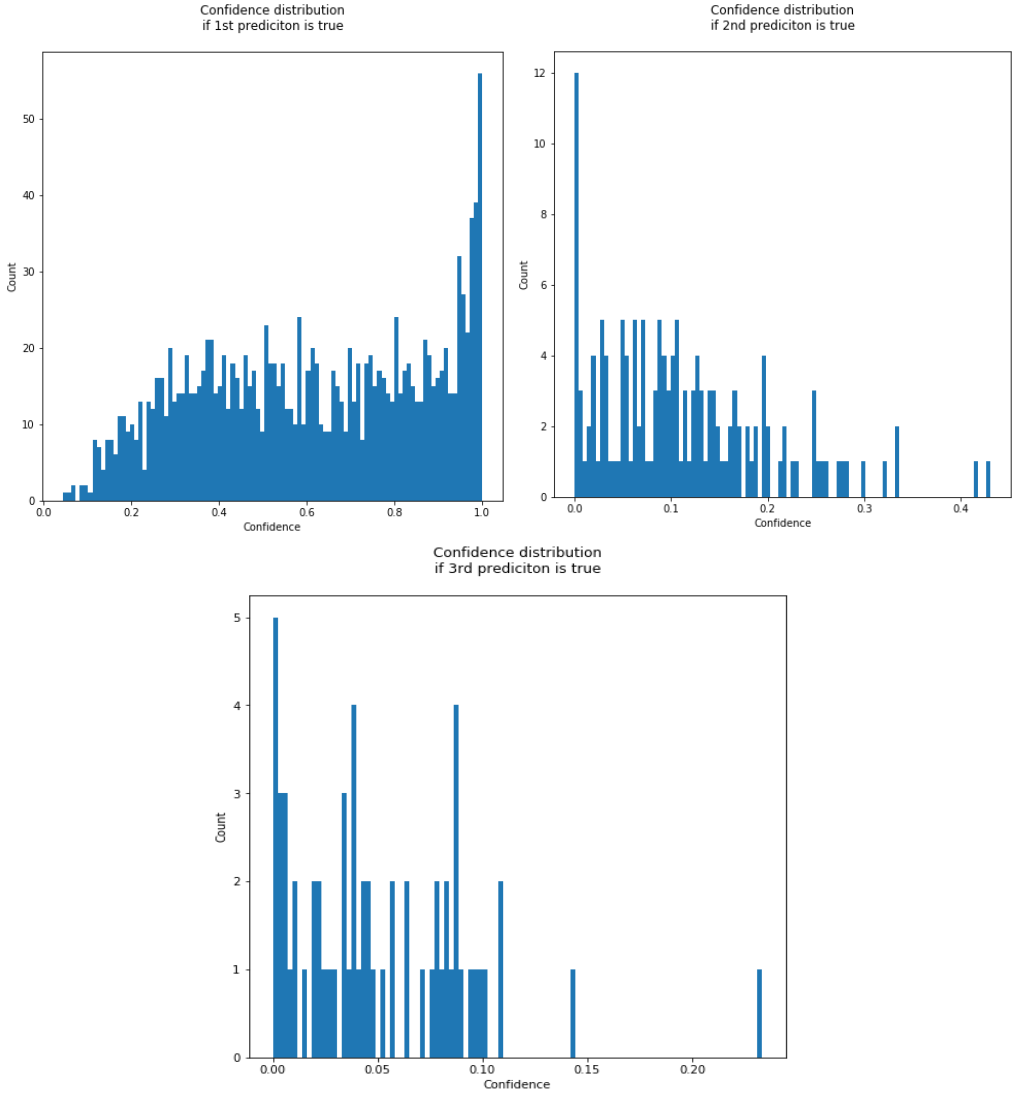
1.Picture. Dog characters and tweets being faovrited.

- With each model prediction (assuming that “guess” was false), the rate of being true is decreasing exponentially.



2.Picture. Prediction rate on each attempt of predicting.

- The confidence rate of each next prediction tend to be lower. Looking to histograms it seems that values are decreasing exponentially. It was expected to have lower values on every next attempt such a decrease of confidence values is a surprise. Further investigations has to be done to check is ti due to quality of pictures or algorithm should be improve.



3.Picture. Distribution of confidence values on each true attempt (assuming that previous one was wrong).

This is just a small amount of what we can find from this data. Overall, WeRateDogs is a beautiful phenomena due to its humor, unique score system, share popularity and memes. Investigation of this data could enlarge our understanding about humans and what makes us love dogs in general.