Insights from Analyzing WeRateDogs Data

WeRateDogs is a twitter account that rate people's dog. The rating is usually in the scale of 0 to 10 but the score can be higher than 10 too. The WeRateDogs dataset includes basic information about the tweets (e.g. tweet id, comment text, rating, timestamp, etc.) as well information about the number of times each tweet has been retweeted or marked as favorited. This data was combined with a separate dataset that used image processing to classify breeds of dogs based on the image provided in the tweets.

Figure 1 shows the top ten breeds of dogs that have the most number of retweets. Golden_retriever and Labrador_retriever are the most retweeted dog breed. Golden_retriever alone has more than fives times retweets than most of other brands.

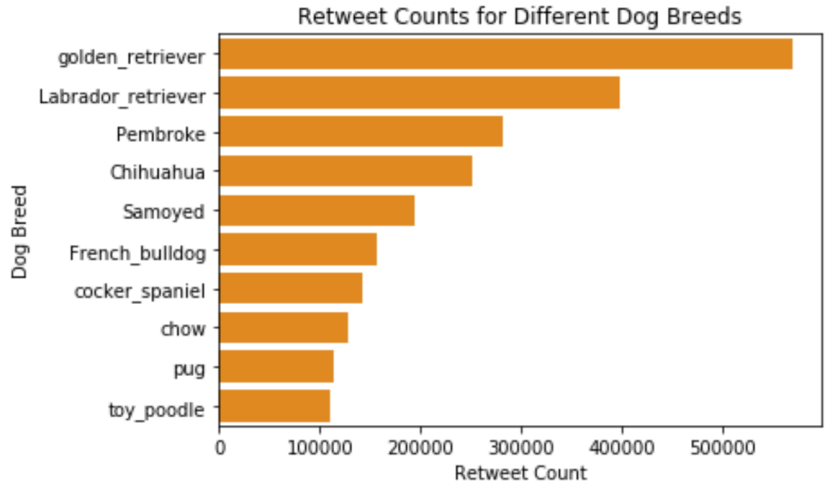

Figure 1-Retweet counts for different dog breeds

Figure 2 shows the most favorite dog breeds. Similar to the results of retweet counts, Golden_retriever and Labrador_retriever are the dog breeds with the most favorite counts. In addition, the top ten dog breeds with the most number of retweets also have the most number of favorite counts.

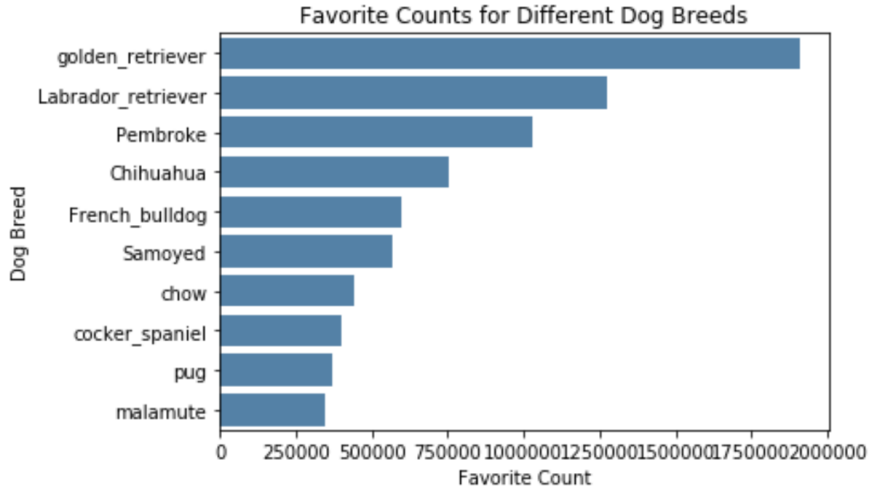

Figure 2-Favorite counts for different dog breeds

Figure 3 shows the three sources that were used to tweet about the dogs (namely iphone, twitter web client, and tweetDeck). The twitter for Iphone has been used the most to send tweets with #WeRateDogs. The web client and TweetDeck are not very popular sources.

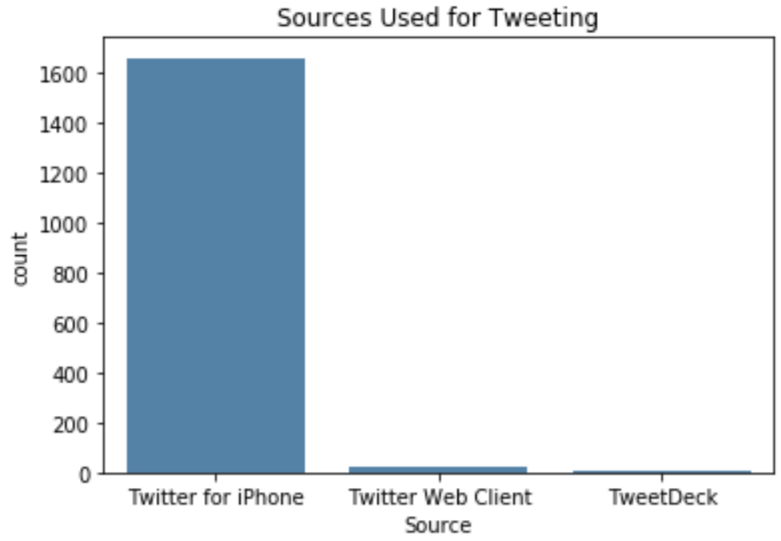

Figure 3-Sources used for tweeting to WeRateDogs

Figure 4 shows the number of times each prediction algorithms was identified as the most probable prediction. This was obtained by comparing the confidence level of each algorithm for each tweet and marking the highest number as the most probable dog breed. The first prediction algorithm has the greatest number of incidents with the maximum confidence level, showing that this algorithm does a better job at identifying images compared to the second and the third algorithms. Prediction 3 algorithm has the poorest outcome by far.

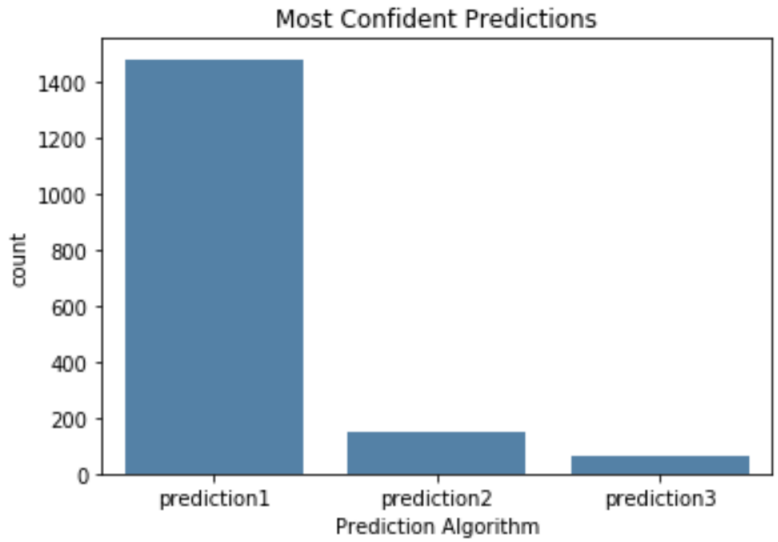

Figure 4-Most confident prediction algorithm