ANALYSE OF THE WERATEDOGS'S TWEETS

Udacity - Nanodegree Data Analyst

Leonardo Simões

07/15/2020

Summary

1. Which tweet got the highest score?

The highest 'rating', which is calculated by 'rating_numerator' / 'rating_denominator', is 177.6, considering only the data obtained and cleared. The dog's name is Atticus, and the post for him has a count of 2772 retweets and 5569 favorites. The id tweet 749981277374128128 is the one with this ‘rating’.

Figure 1 – Atticus in WeRateDogs’s Twitter

The image recognition algorithms p1, p2 and p3 applied to the image did not recognize him as a dog, but recognized two other elements present in the image, sunglasses and a bow tie.

2. How are the measurements of each stage?

There is a big difference in the number of dogs per stage. 84% of the stages are 'none'. Considering those that are classified (there is no stage 'none') 66% are 'pupper', 20% are doggo, 7.5% are 'puppo', 2.3% are 'floofer' and and the rest is classified as with 2 stages.

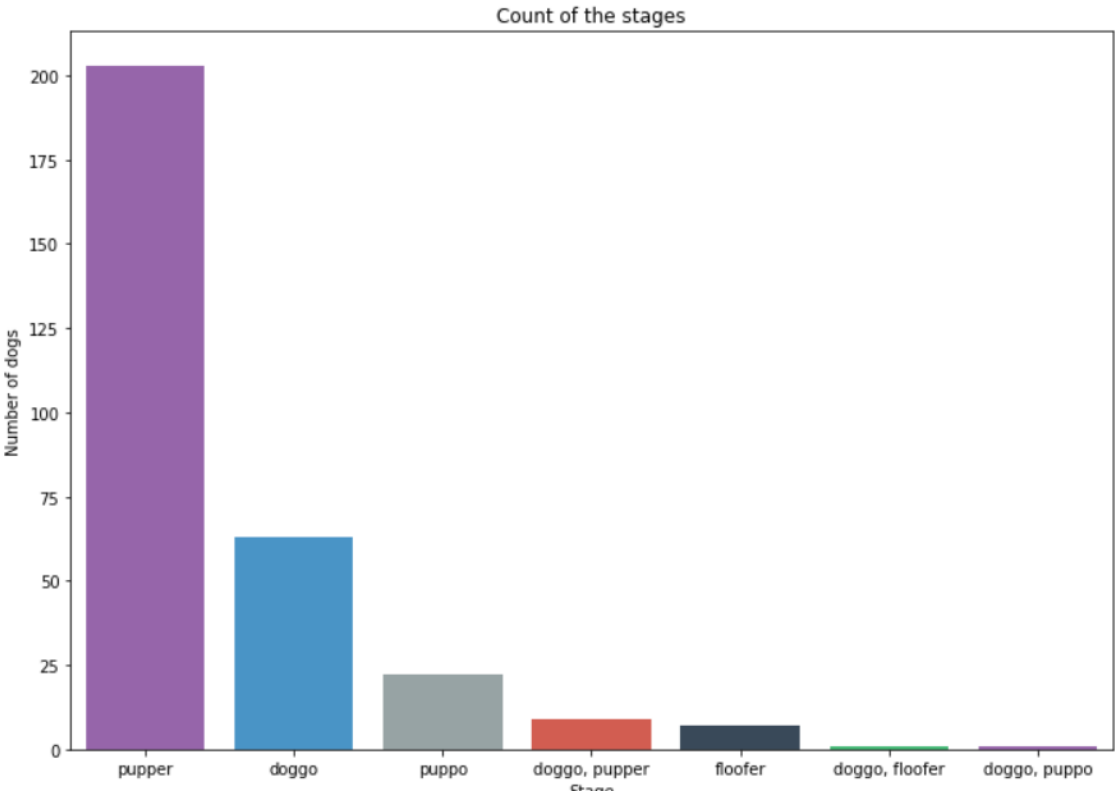

Figure 2 – Count of the stages

stages	rating	rating_numerator	rating_denominator	img_num	p1_conf	p2_conf	p3_conf	retweet_count	favorite_count
doggo	1.188889	11.888889	10.000000	1.285714	0.577058	0.157341	0.062083	7125.698413	19356.380952
doggo, floofer	1.100000	11.000000	10.000000	1.000000	0.354733	0.177538	0.131706	3433.000000	17169.000000
doggo, pupper	1.100000	11.000000	10.000000	1.111111	0.833500	0.082893	0.027889	4083.444444	12533.111111
doggo, puppo	1.300000	13.000000	10.000000	1.000000	0.321676	0.115138	0.096100	19196.000000	47844.000000
floofer	1.200000	12.000000	10.000000	1.142857	0.609204	0.155319	0.047997	4968.714286	13206.000000
none	1.180085	12.495853	10.628555	1.191943	0.590568	0.134187	0.060981	2577.957346	8478.924763
pupper	1.071429	10.714286	10.000000	1.251232	0.604692	0.130573	0.056709	2363.581281	7197.738916
puppo	1.200000	12.000000	10.000000	1.454545	0.722293	0.135468	0.043701	6473.954545	21582.090909

Figure 3 - Average values for each stage

Considering only the unique stages, the ‘puppo’ stage has the highest average favorite count, followed by ‘doggo’, ‘floofer’ and ‘pupper’. Considering only the unique stages, the ‘doggo’ stage has the highest average retweet count, followed by ‘puppo’, ‘floofer’ and ‘pupper’. The average 'rating' does not differ much from one stage to another, less than 0.3.

3. How are the results of the algorithms p1, p2 and p3 compared to each other?

The p2 algorithm was the one that identified the largest number of dogs, 1553. Followed by p1 with 1532 and p3 with 1499. This metric slightly differentiated the efficiency of the algorithms, but the difference between them was not very large.

The p1 algorithm is the one with the highest average confidence, around 59%. The p2 and p3 algorithms have an average confidence of approximately 13% and 6%, respectively. This metric was able to quite distinguish the supposed efficiency of the algorithms, p1 being reasonably efficient, and p2 and p3 relatively inefficient.

The algorithms p1, p2 and p3, identified 111,113 and 116 different breeds of dogs respectively. In this metric, the algorithms do not differ much.

REFERENCES

UDACITY - Data Analyst Nanodegree Program: https://www.udacity.com/course/data-analyst-nanodegree--nd002

WeRateDogs, Twitter profile (@dog_rates): https://twitter.com/dog_rates/status/749981277374128128