
WeRateDogs 推特数据分析报告

WeRateDogs 是一个推特主，他以诙谐幽默的方式对人们的宠物狗评分。这些评分通常以 10 作为分母。但是分子则一般大于 10：11/10、12/10、13/10 等等。为什么会有这样的评分？因为 “They’re good dogs Brent.” 也有一部分的原因是一张图像里有多只狗：165/170。有时候也有图 1 中的极端情况。

图 1. WeRateDogs 的推特图片

This is Atticus. He's quite simply America af.
1776/10



11:00 PM - 4 Jul 2016

通过分析 WeRateDogs (@dog_rates) 的 1994 条推特数据，我们得到一些有趣的结论。

1 评分

根据推文提取到 1577 个有效评分，标准评分（评分分子*10/评分分母）的均值为 10.68，大部分评分都在满分 10 分以上，最多次数的评分为 12 分，其次为 11 分。

那分数特别低的是些什么狗呢？Emm，多半是比较可爱的其他动物啦（见图 3-6）。

图 2. 标准评分直方图

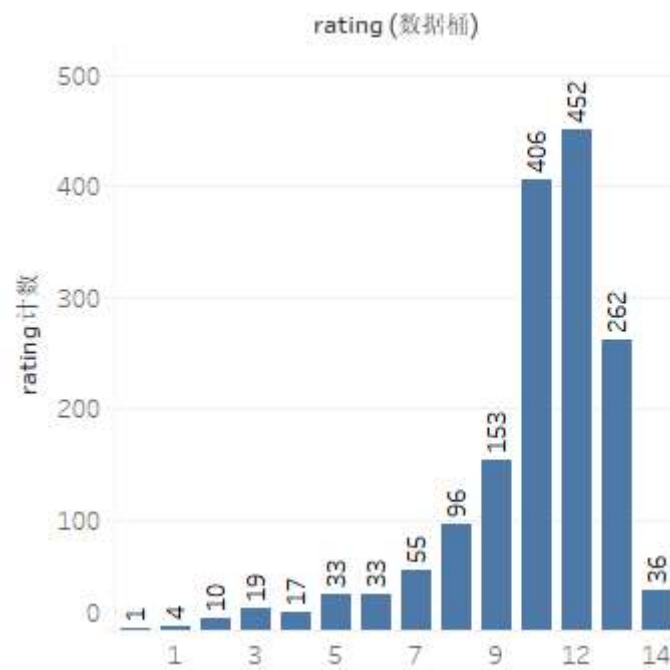


图 3-6. WeRateDogs 的推特图片

This is a funny dog. Weird toes. Won't come down. Loves branch. Refuses to eat his food. Hard to cuddle with. 3/10



Not familiar with this breed. No tail (weird). Only 2 legs. Doesn't bark. Surprisingly quick. Shits eggs. 1/10



Unfamiliar with this breed. Ears pointy af. Won't let go of seashell. Won't eat kibble. Not very fast. Bad dog 2/10

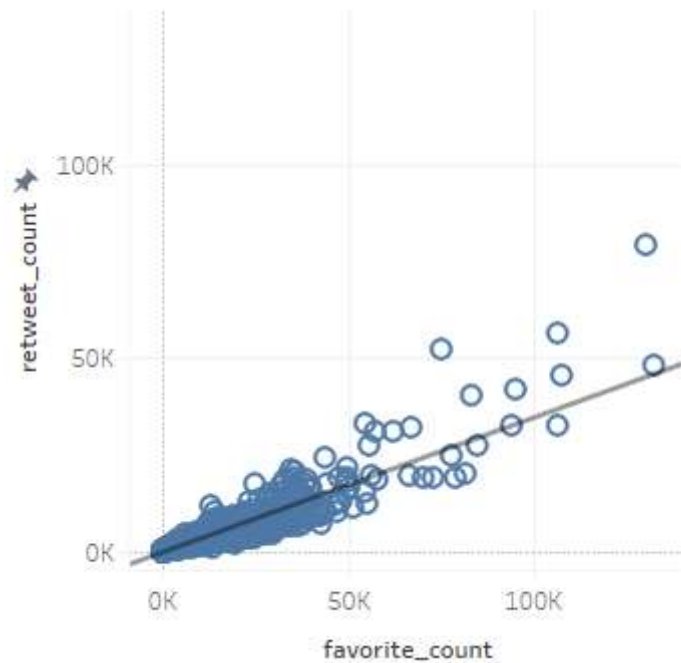


This is Charl. He's a bully. Chucks that dumbbell around like its nothing. Sharp neck. Exceptionally unfluffy. 3/10



2 喜爱数和转发数

图 7. 喜爱数和转发数的关系图



平均喜爱数为 8923，平均转发数为 2770，两者呈线性相关。

3 狗狗地位对评分、喜爱数和转发数的影响

Puppo、doggo 和 floofer 的评分都在 12 分以上，而且差距并不大。

puppo 平均收到 22644 次喜爱，是平均收到最多次喜爱的狗狗地位。但平均被转发次数最多的狗狗地位却是 doggo，doggo 平均被 7088 次转发。puppo 平均被转发次数紧随其后，为 6986 次。

图 8. 狗狗地位对评分、喜爱数、转发数的影响



4 神经网络对狗狗种类的预测

通过一个可以对狗狗种类进行分类的神经网络，对 WeRateDogs 推特档案中的所有图像进行预测，发现判断为狗的平均预测可信度为 64.44%，判断为非狗的平均预测可信度为 53.56%，两者差距并不算大，看来神经网络还要再多多加强机器学习呀。

而 Golden retriever 和 labrador retriever 分别是被神经网络识别出次数最多和第二多的狗狗种类。可能 retriever 是最经常被养的宠物狗，或者是最容易被神经网络识别出来的狗狗品种。

图 9. 神经网络对狗狗种类预测的可信度

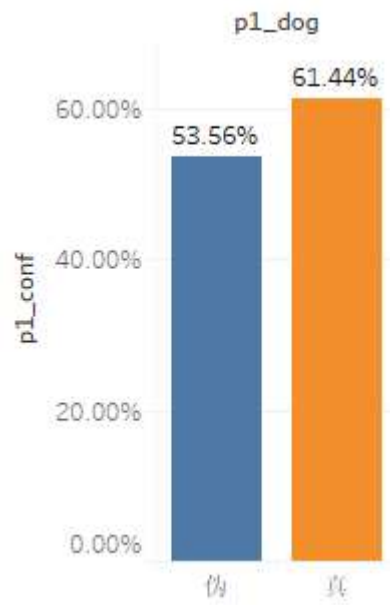


图 10. 被识别出最多次数的狗

