

Project overview and path forward

Factors that affect pet adoption speed

Tim from petfinder.com



Why pet adoption?



The Addams
Family Litter from
petfinder.com

Animals are awesome and all of our group either likes animals or has pets. Finding a way to predict adoption speed and possibly finding what features contribute to the speed at which different pets are adopted to help shelters plan better for the animals care and/or maybe helping them to see what features might make a pet more adoptable -- this model is a way to help these fluffy-balls we all like have better lives in the shelter, and possibly to get out of the shelter more quickly to be with their potential people sooner

Data Description



Baxter(adopted)
from petfinder.com

Data pulled from a kaggle 2018 competition:
PetFinder.my Adoption Prediction

Our data contains thousands of entries of pets, each of which contains information including

- Type - dog or cat
- Name - if given one
- Quantity - some come as a set of siblings, some are mama's with babies, most individuals
- Physical descriptors: color, breed, fur length, health, shots, sterilized, maturity size and age
- number of pictures or videos posted
- a written description of the pet(s)
- Fee from free and up
- RescuerID & State - who found, and where the pet is
- Adoption Speed - how long it took them to be adopted, & what we are trying to create a model to predict

How is the data going to answer our question

We will need to be able to predict the outcome of adoption speeds of the pets. Once we have that model we can make other inferences based on the largest factors/features that play into adoption speed prediction. As we have the adoption speed already we can compare our model to actual results



Wren from
[petfinder.com](https://www.petfinder.com)

Data Exploration



Calvin from [petfinder.com](https://www.petfinder.com)

14993 rows, 24 columns

We found 12 nulls in description, we dropped those entries as the number of nulls was low, and as we wanted to use description word count in the model. So we changed description to a length, as it is something measurable we can include into the model

Binned quantity of pets, fee & photo amount into categories as they each had too many unique values and we wanted to scale the data

Several columns where ID columns - will not use in model as they are just 'noisy' data

Description of the analysis phase



Luca from
petfinder.com

Supervised Machine Learning Model to predict adoption speed. Took the cleaned/transformed data from exploration and use.

We have 14981 data points with **5 possible outcomes** for adoption speed split 75%/25% into train and test data.

- 0 - adopted the same day of listing
- 1 - adopted on the first week of listing
- 2 - adopted on the first month of listing (7-30 days)
- 3 - adopted on the 2nd or 3rd month following the listing
- 4 - stayed over 90 days of listing

Target Accuracy 42.3+% as that would at least get us in the 'bronze'. The top 10 (non-cheating) finalists were all at 44.1+% in the initial contest 3 years ago

Assumptions

- We expected the type of pet to have an impact on our model
- Pet color was also expected to have a large impact, especially the first color
- The health of the pet was expected to have a significant impact
- We the age of the pet to have a large impact. Everyone loves kittens and puppies
- The amount of photos of the pet was also expected to have an impact

RandomForest Models

We used a standard scaler prior to the train test split to standardize the data for the model

We have used feature importance pulled from first model to improve the models following

75/25 train test split

First model run all features: accuracy 41.4%

Feature importance (add image?)

(image of where that puts us in 'competition'?)

Comp ranking: around 632/633 out of 2023 entries, not in bronze though

RF Models contin.

Second model run less type, health and video amount: accuracy 41.6% best model created in RF at 75/25 split (add image)

Comp ranking: around 621/622 out of 2023 entries, not in bronze though

Ran several other RF iterations 75/25 - dropping more features, changing word count feature, etc. - *but none produced as strong of model as the first two passes*

RF Models contin.

Adjusted split to 80/20 and use best two models from first iterations

First model run all features: accuracy 42.2%

Feature importance (add image?) - changed slightly from first model age and photo amount switched places

Comp ranking: 245/246 out of 2023 entries, not in bronze though but only barely

Second model run less type, health and video amount: accuracy 42.5% best model (so far)

Comp ranking: 141/142 out of 2023 entries, **in bronze!**
Min target accuracy goal hit

Findings



Bonnie from petfinder.com

Top 3 features

- Word count
- Age
- Photo amount

Features that didn't matter/ held back model

- Type of pet
- Health
- Video Amount

Some of these results were surprising to us, especially the word count feature. Once we dropped the bottom three features, our models improved.

If had time---

Will add as the end of the project approaches

Thank you!

Fluff from petfinder.com



Pet sources

Tim

Addams Family Litter

Baxter(adopted)

Wren

Calvin

Luca

Bonnie