Problem Statement

As an animal lover, I chose the PetFinder.my - Pawpularity Contest as my capstone project. It's a $25000 prize ongoing competition on kaggle. PetFinder.my is Malaysia's leading animal welfare platform, featuring over 180,000 animals with 54,000 happily adopted. But there are still millions of stray animals suffer on the streets or are euthanized in shelters every day around the world. We hope that pets with attractive photos can generate more interest and be adopted faster with the help of data science. The purpose of the project is to accurately determine a pet photo's appeal and even suggest improvements to give these rescue animals a higher chance of loving homes.

Dataset Information

This dataset has both image data and csv file which store some basic information about the image. There are 9912 image data, with 1.06GB in size, are stored in a jpg image format in training folder and 12 basic features for each photo and the photo's Pawpularity score are in the train.csv file where the Id column gives the photo's unique Pet Profile ID corresponding the photo's file name. The Pawpularity score are continuous number between 0 and 100. These 12 features (Focus, Eyes, Face, Near, Action, Accessory, Group, Collage, Human, Occlusion, Info and Blur) are labeled with the value of 1 (Yes) or 0 (No).

Expected Outcome

Given a photo of a pet animal and some basic information about the photo, we should be able to estimate the 'pawpularity' score of the pet.

Evaluation

Submissions are evaluated on Root Mean Squared Error (RMSE) between the predicted value and the observed target.

Preliminary Result

Treat the problem as a regression problem and use Random Forest regression on the metadata. After tuning the hyperparameters of Random Forest regression, I got a test RMSE with 20.64.

Anticipated Challenges

The big challenger for me is to use both image data and metadata to do the modeling.

Plan for next steps

Model building on image data.

Model building on both image data and metadata.