The Data Incubator
Capstone Project Proposal

Rong Wang

11/18/2021

Research Code Competition

PetFinder.my

PetFinder.my - Pawpularity Contest

Predict the popularity of shelter pet photos

$25,000

Prize Money

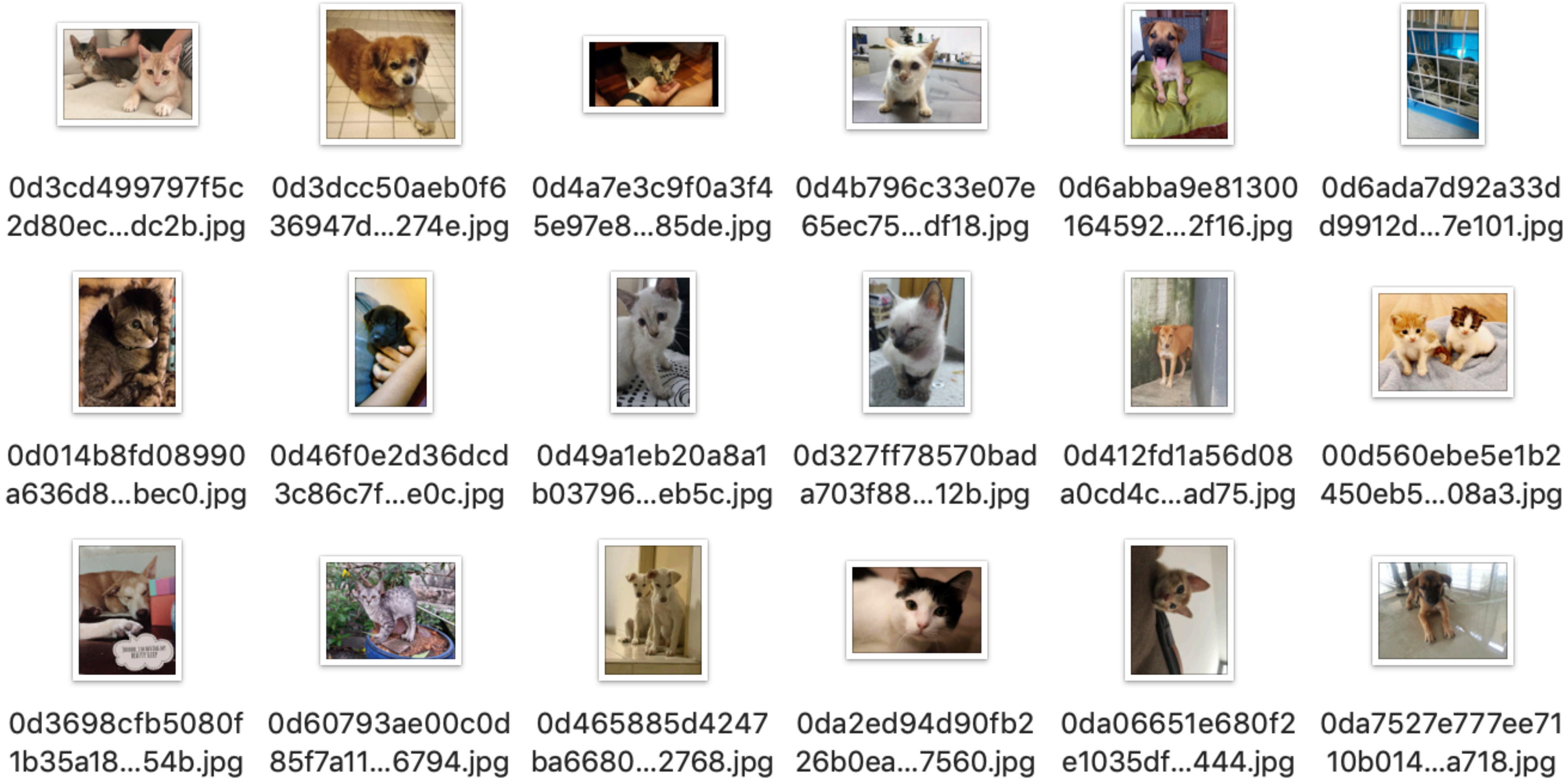
PetFinder.my · 2,046 teams · 2 months to go (2 months to go until merger deadline)

PetFinder.my - Pawpularity Contest

- A $25,000 prize ongoing competition on Kaggle
- Predict the popularity of shelter pet photos

Dataset Information

- 9912 image data, 1.06GB in size



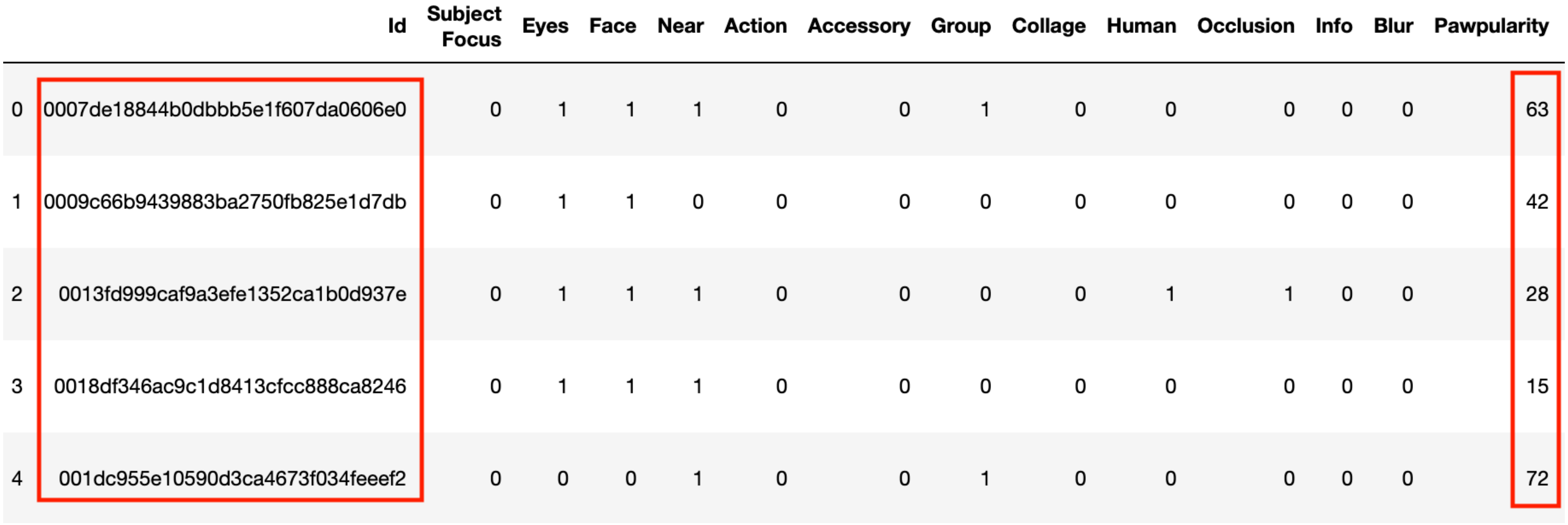
0d3cd499797f5c 2d80ec...dc2b.jpg
0d3dcc50aeb0f6 36947d...274e.jpg
0d4a7e3c9f0a3f4 5e97e8...85de.jpg
0d4b796c33e07e 65ec75...df18.jpg
0d6abba9e81300 164592...2f16.jpg
0d6ada7d92a33d d9912d...7e101.jpg

0d014b8fd08990 a636d8...bec0.jpg
0d46f0e2d36dcd 3c86c7f...e0c.jpg
0d49a1eb20a8a1 b03796...eb5c.jpg
0d327ff78570bad a703f88...12b.jpg
0d412fd1a56d08 a0cd4c...ad75.jpg
00d560ebe5e1b2 450eb5...08a3.jpg

0d3698cfb5080f 1b35a18...54b.jpg
0d60793ae00c0d 85f7a11...6794.jpg
0d465885d4247 ba6680...2768.jpg
0da2ed94d90fb2 26b0ea...7560.jpg
0da06651e680f2 e1035df...444.jpg
0da7527e777ee71 10b014...a718.jpg

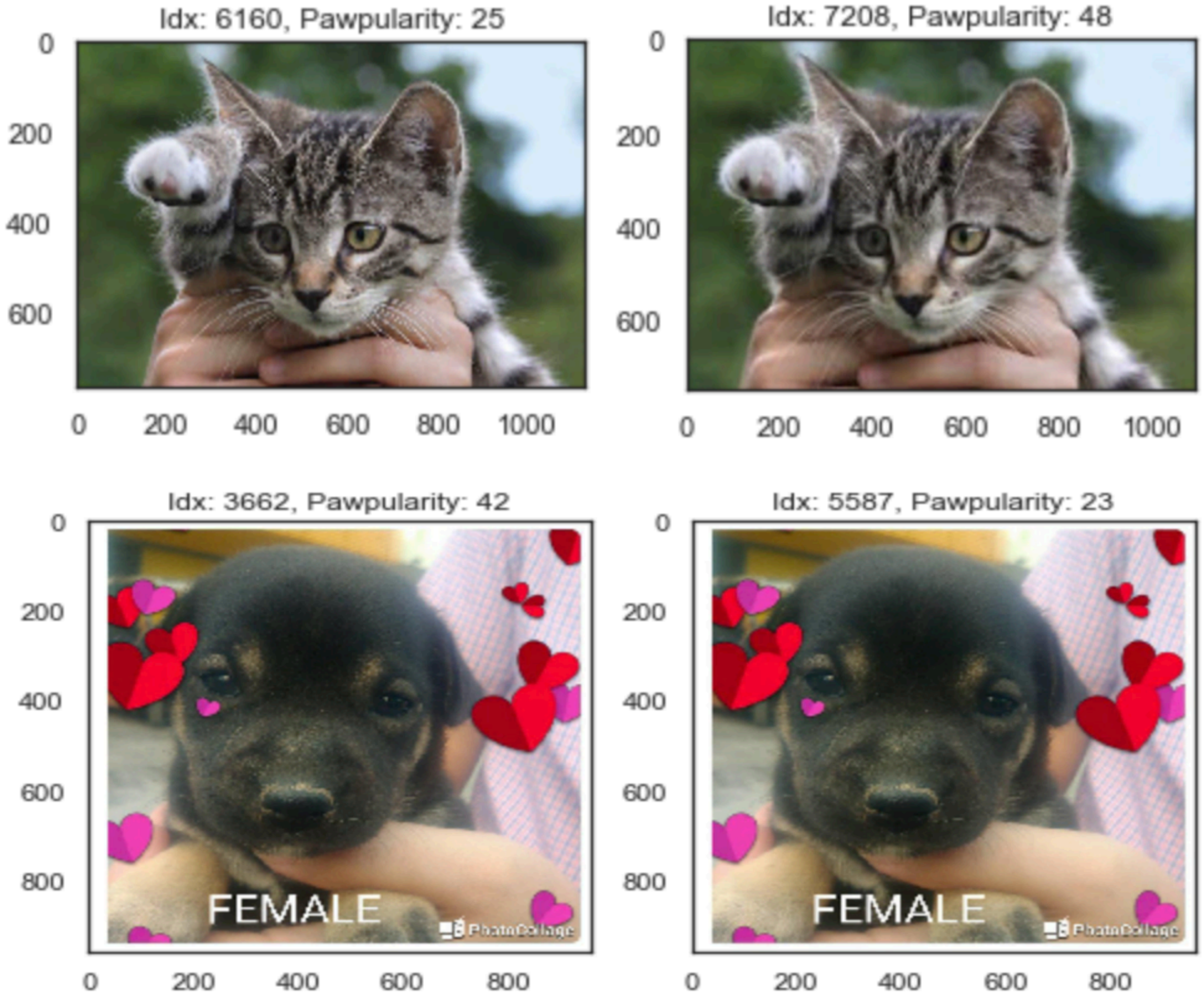
Dataset Information

- Metadata (train.csv)



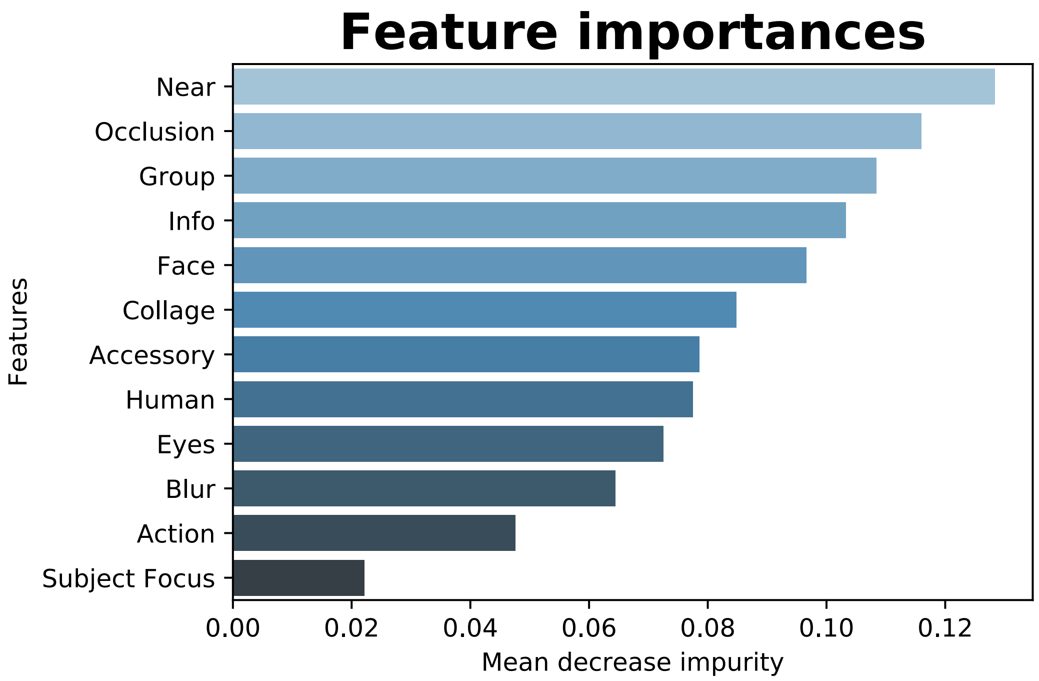
	Id	Subject Focus	Eyes	Face	Near	Action	Accessory	Group	Collage	Human	Occlusion	Info	Blur	Pawpularity
0	0007de18844b0dbbb5e1f607da0606e0	0	1	1	1	0	0	1	0	0	0	0	0	63
1	0009c66b9439883ba2750fb825e1d7db	0	1	1	0	0	0	0	0	0	0	0	0	42
2	0013fd999caf9a3efe1352ca1b0d937e	0	1	1	1	0	0	0	0	1	1	0	0	28
3	0018df346ac9c1d8413cfcc888ca8246	0	1	1	1	0	0	0	0	0	0	0	0	15
4	001dc955e10590d3ca4673f034feeef2	0	0	0	1	0	0	1	0	0	0	0	0	72

Data Preprocessing: removed 27 pairs of duplicate images



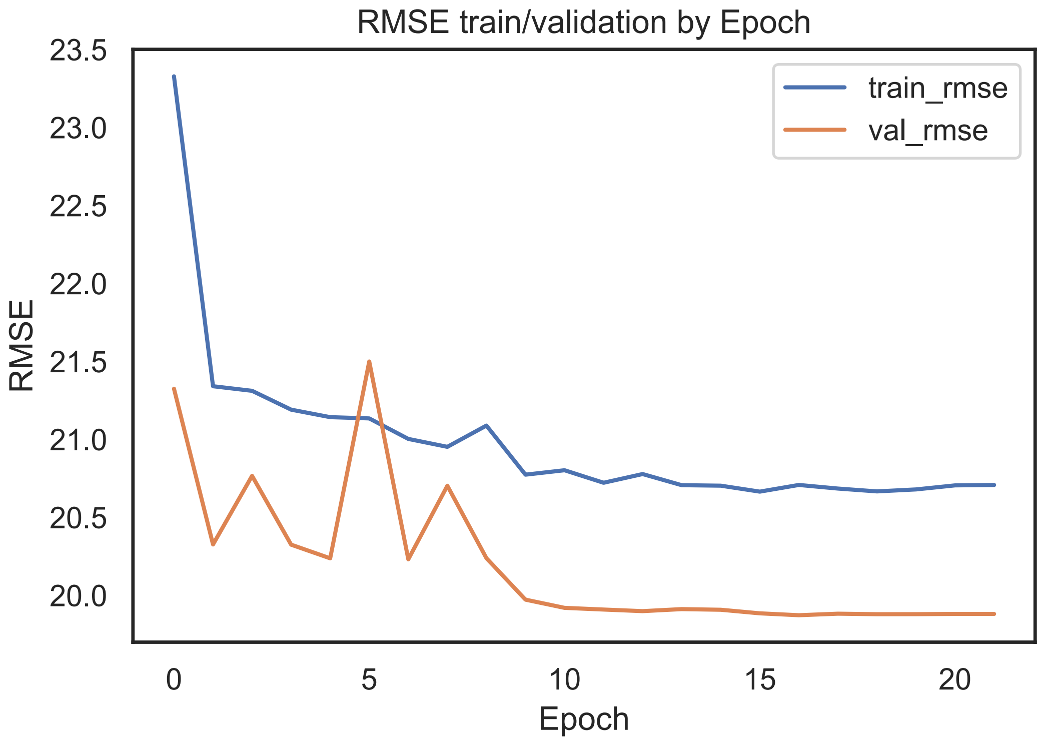
Preliminary result:
Random forest regression on metadata

- Tuned three parameters: *bootstrap, max_features, number of trees and max_depth.*
- The best values for these four parameters are: *True, log2, 400, 10*.
- Train RMSE: **20.4**.
- Validation RMSE: **20.74**.



Preliminary result: CNN on image data

- Basic CNN model
- Pretrained model(EfficientNetB0.h5)
- Train RMSE stabilized at **20.68**.
- Validation RMSE stabilized at **19.88**.



Thank you!