

PET ADOPTION PREDICTION

- SETH AGEE
- BRELYN GRANT
- BRENDAN ROBERTS
- RILEY ROBINSON



INTRO

- Data Set: Kaggle PetFinder database
 - Problem: Bad adoption listings can prevent pets from being adopted quickly or at all.
 - Goal: Create an application that asks the user for details on their pet's adoption listing, calculates predicted adoption speed, and offers suggestions to improve the listing.
- Adoption Speeds:
 - 0 = 1 day
 - 1 = 1 week
 - 2 = 1-4 weeks
 - 3 = 1-3 months
 - 4 = >3 months

DATA CLEANING AND TRANSFORMATION

Initial Data Set

1	Type	Name	Age	Breed1	Breed2	Gender	Color1	Color2	Color3	MaturityS	FurLength	Vaccinate	Deworme	Sterilized	Health	Quantity	Fee	State	RescuerID	VideoAmt	Descriptio	PetID	PhotoAmt	AdoptionSpeed
2	2	Nibble	3	299	0	1	1	7	0	1	1	2	2	2	1	1	100	41326	8480853f	0	Nibble is a	86e1089a	1	2
3	2	No Name	1	265	0	1	1	2	0	2	2	3	3	3	1	1	0	41401	3082c712	0	I just four	6296e909	2	0
4	1	Brisco	1	307	0	1	2	7	0	2	2	1	1	2	1	1	0	41326	fa90fa5b1	0	Their preg	3422e490	7	3
5	1	Miko	4	307	0	2	1	2	0	2	1	1	1	2	1	1	150	41401	9238e4f4	0	Good gua	5842f1ff5	8	2
6	1	Hunter	1	307	0	1	1	0	0	2	1	2	2	2	1	1	0	41326	95481e95	0	This hand	850a43f9	3	2
7	2		3	266	0	2	5	6	0	2	1	2	2	2	1	1	0	41326	22fe332b	0	This is a st	d24c30b4	2	2
8	2	BULAT	12	264	264	1	1	0	0	2	3	2	2	3	1	1	300	41326	1e0b5a45	0	anyone wi	1caa6fcd	3	1
9	1	Siu Pak &	0	307	0	2	1	2	7	2	1	2	2	2	1	6	0	41326	1fba5f6e5	0	Siu Pak ju	97aa9eea	9	3
10	2		2	265	0	2	6	0	0	2	2	2	2	2	1	1	0	41326	d8af7afec	0	healthy ar	c06d167c	6	1

Data Transformation

M	N	O
Vaccinated	Dewormed	Sterilized
2	2	2
1.5	1.5	1.5
1	1	2
1	1	2
2	2	2
2	2	2
2	2	1.5
2	2	2
2	2	2
1.5	1.5	1.5
2	1	2

Derived Attributes

AL	AM	AN	AO	AP	AQ	AR	AS
Num_Colors	Is_Mutt	Has_Name	Description_Length	Has_Picture	Has_Fee	Multi_Listing	AdoptionWeeks
3	1	1	723	1	0	1	1
1	0	1	26	1	0	0	14.3
1	0	1	19	1	0	0	14.3
1	0	1	127	1	0	1	8.4
1	1	0	193	1	0	0	0
2	0	1	561	1	0	0	14.3
3	1	0	188	1	0	1	14.3
2	1	1	170	1	0	0	8.4
2	0	1	758	1	0	0	8.4
2	1	0	343	1	0	1	14.3
1	0	1	518	1	0	0	3.1
2	0	0	21	1	0	0	14.3

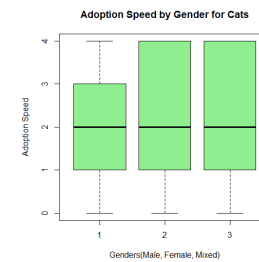
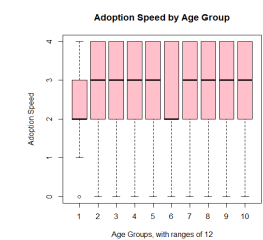
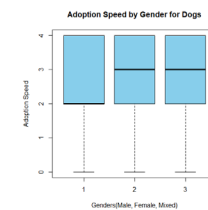
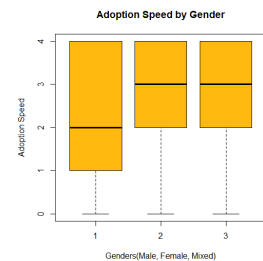
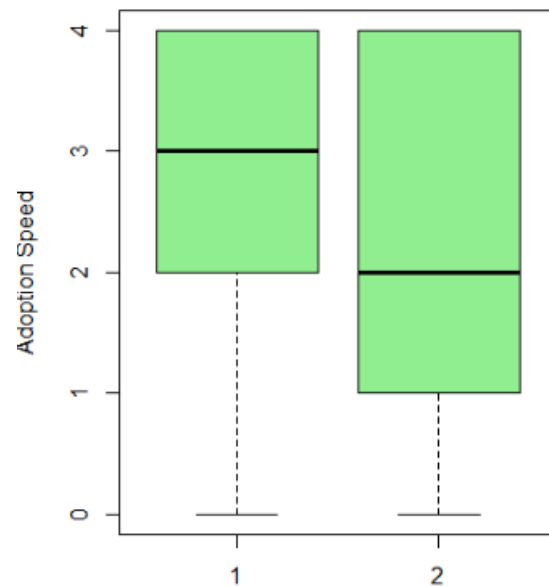
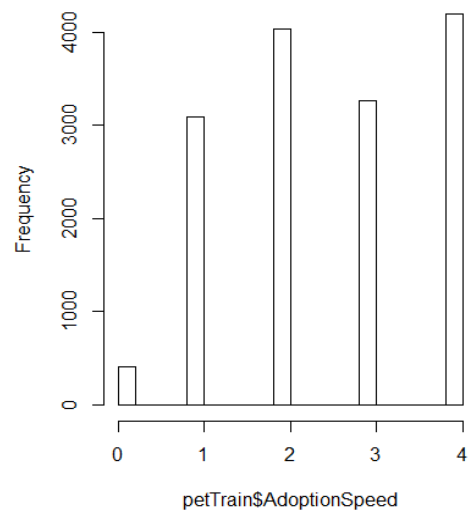
DATA CLEANING AND TRANSFORMATION

	State	State_Avg_Speed	State_Median_Speed	State_Listing_Count
1	41326	2.452031214	2	8714
2	41401	2.540182055	3	3845
3	41327	2.78173191	3	843
4	41336	2.41025641	2	507
5	41330	2.821428571	3	420
6	41332	2.845849802	3	253
7	41324	3.065693431	4	137
8	41325	2.681818182	3	110
9	41335	2.317647059	2	85
10	41361	2.576923077	3	26
11	41345	2.545454545	3	22
12	41367	2.4	2	15
13	41342	3.307692308	4	13
14	41415	3	3	3
15	41380	2	2	0

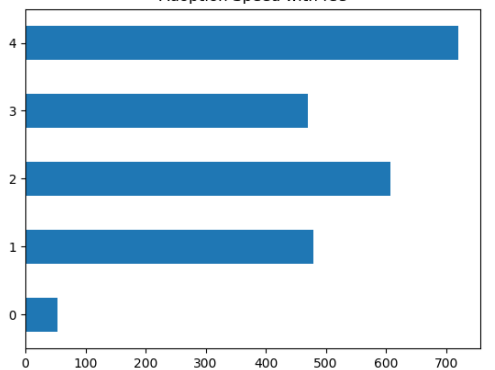
[illegible]

EXPLORATORY DATA ANALYSIS

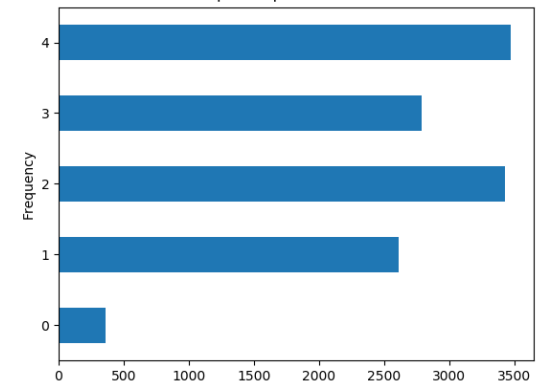
Histogram of petTrain\$AdoptionSpeed



Adoption Speed with fee



Adoption Speed with no fee



TECHNIQUES USED

Rounding was used in all regressions to create a classification problem

Hyperparameters were chosen by stepping through a range possible values

Competitive Models:

- Multiple Linear Regression
- Lasso Regression
- Double Lasso Regression
- Lasso Regression with Thresholds

Not-So Honorable Mentions:

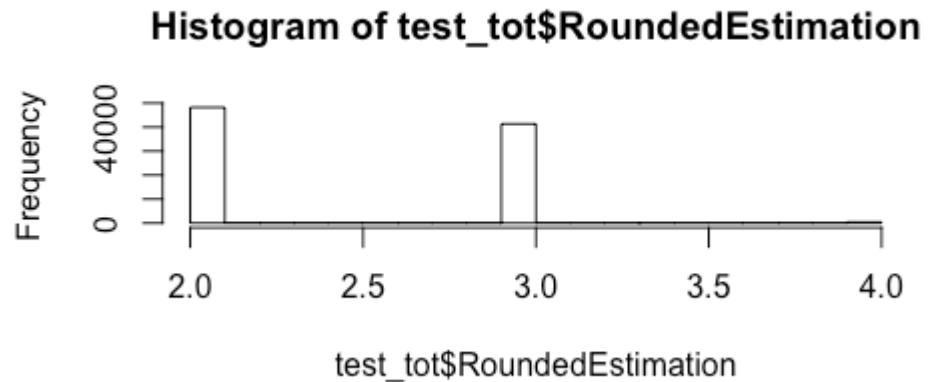
- Subset-Based Multiple Linear Regression
- KNN

Each model was selected by outperforming other models of the same type. The models were run 30 times to find average prediction accuracy.

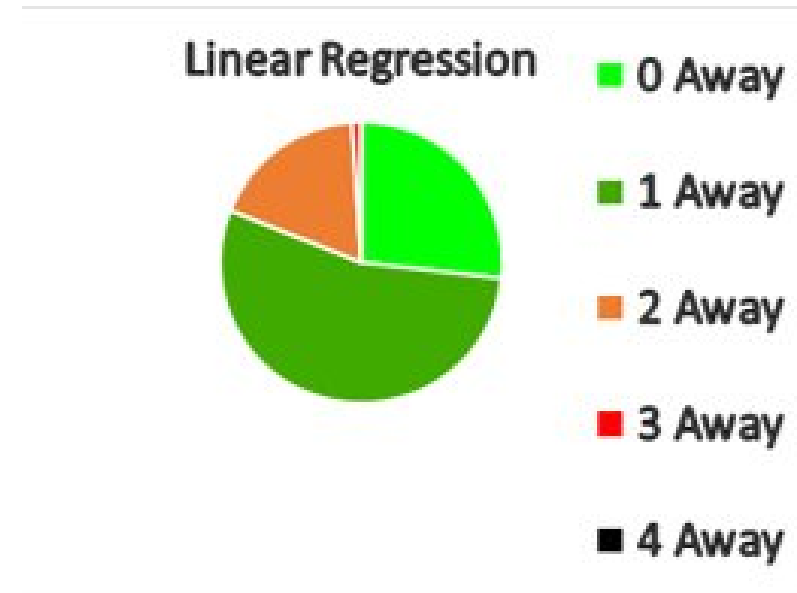
LINEAR REGRESSION

Correct: 27%
One-Off: 54%
Correct and One-Off: 81%

Histogram of Predictions



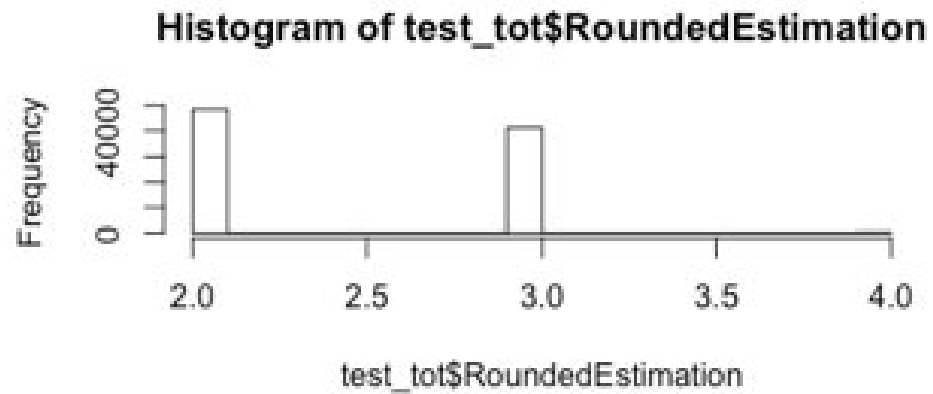
Distance From True Adoption Speed



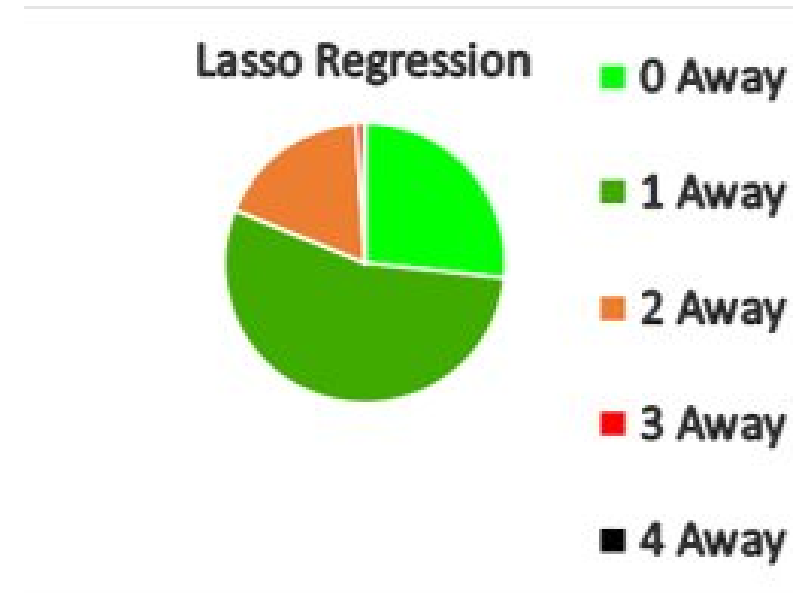
LASSO REGRESSION

Correct: 27%
One-Off: 54%
Correct and One-Off: 81%

Histogram of Predictions



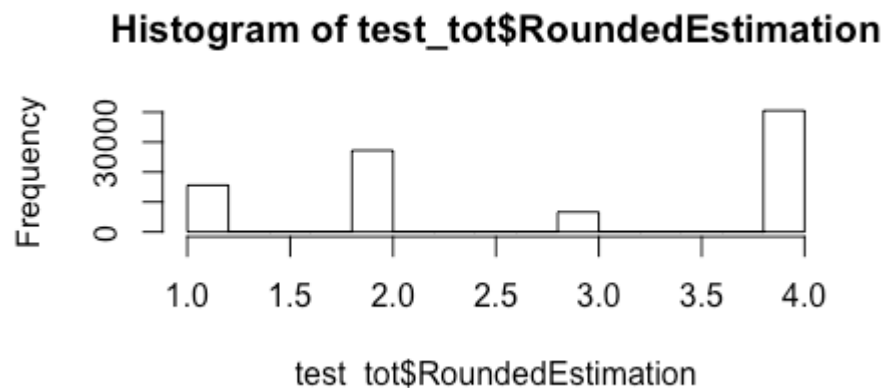
Distance From True Adoption Speed



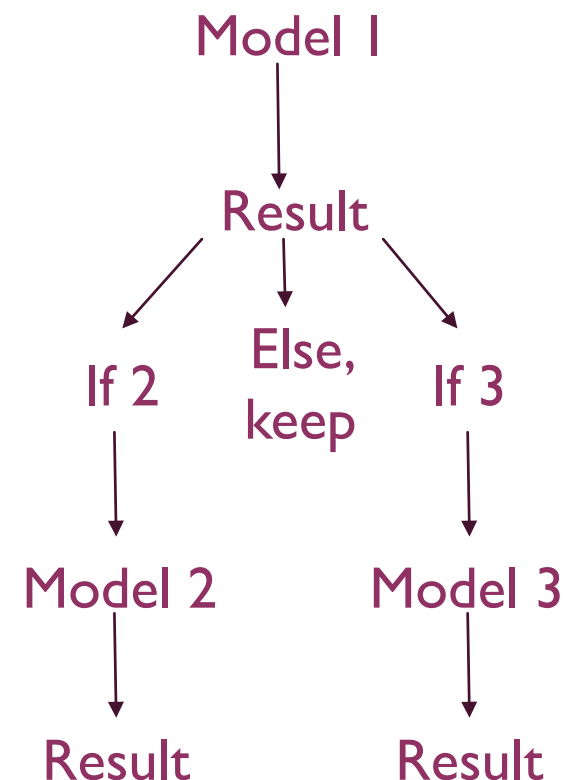
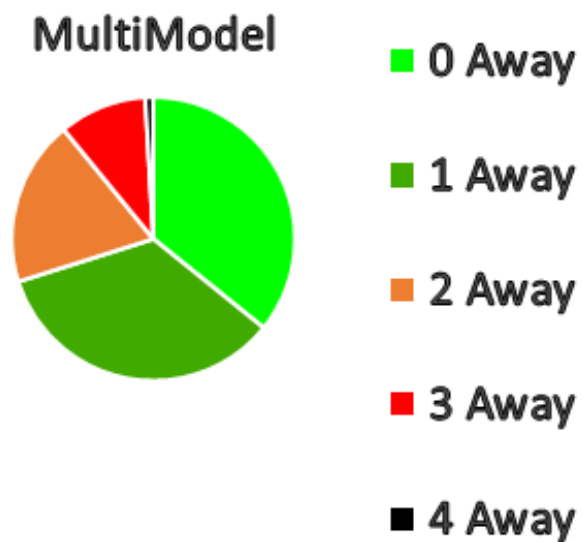
DOUBLE LASSO REGRESSION

Correct: 36%
One-Off: 34%
Correct and One-Off: 70%

Histogram of Predictions



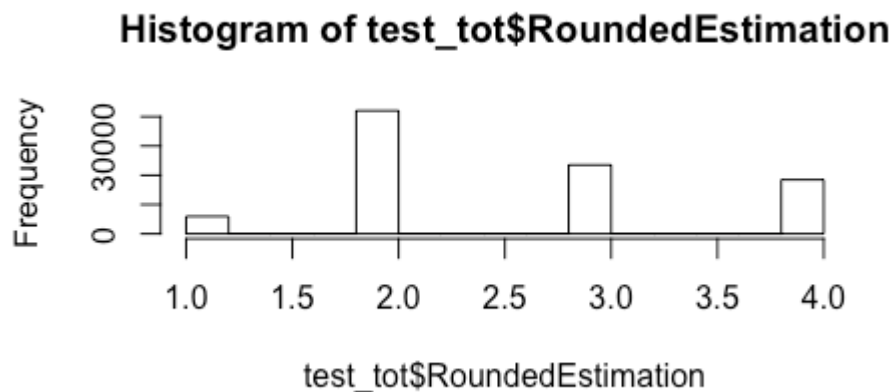
Distance from True Adoption Speed



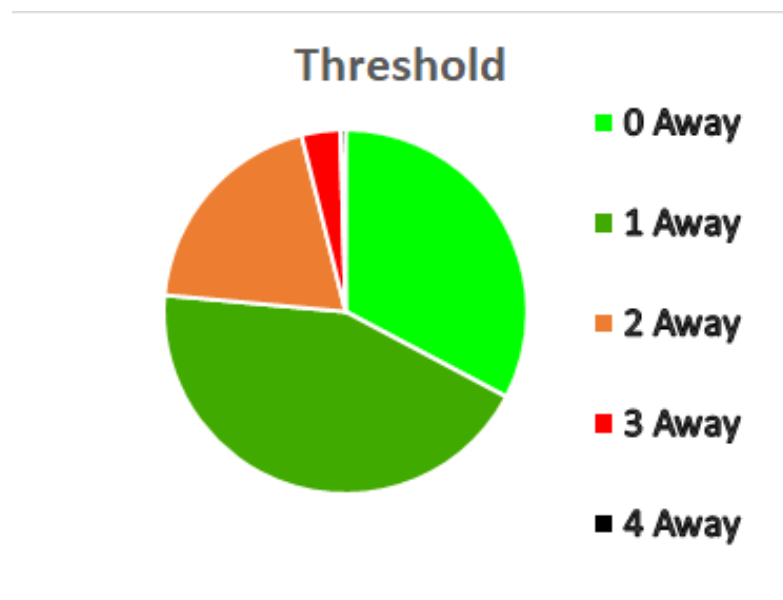
LASSO REGRESSION WITH THRESHOLDS

Correct: 33%
One-Off: 44%
Correct and One-Off: 77%

Histogram of Predictions



Distance from True Adoption Speed

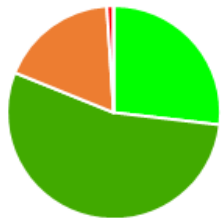


Threshold 1: 2.1

Threshold 2: 2.5

Threshold 3: 2.8

Linear Regression



- 0 Away
- 1 Away
- 2 Away
- 3 Away
- 4 Away

Correct: 27%
One-Off: 54%
Correct or One-Off: 81%

Lasso Regression



- 0 Away
- 1 Away
- 2 Away
- 3 Away
- 4 Away

Correct: 27%
One-Off: 54%
Correct or One-Off: 81%

MultiModel



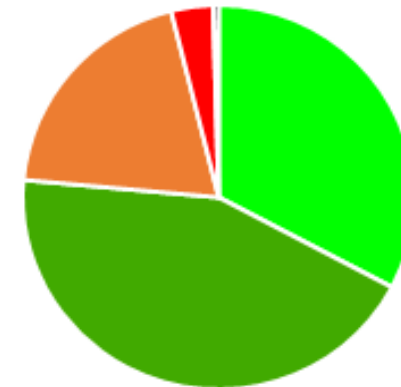
- 0 Away
- 1 Away
- 2 Away
- 3 Away
- 4 Away

Correct: 36%
One-Off: 34%
Correct or One-Off: 70%

Lasso Regression with Threshold was chosen as our model since it was the best all around.

Correct: 33%
One-Off: 44%
Correct or One-Off: 77%

Threshold



- 0 Away
- 1 Away
- 2 Away
- 3 Away
- 4 Away

APP DEMONSTRATION

- Demo Pet
 - Species: Cat
 - Name: Puff
 - Age: 1 month
 - Gender: Male
 - Primary color: White
 - Maturity Size: Small
 - Fur length: Long
 - Unsure about vaccinated, dewormed, sterilized
 - Healthy
 - No fee
 - No pics
 - No videos
 - Is a mutt
 - 30 previous listings
 - In Labuan