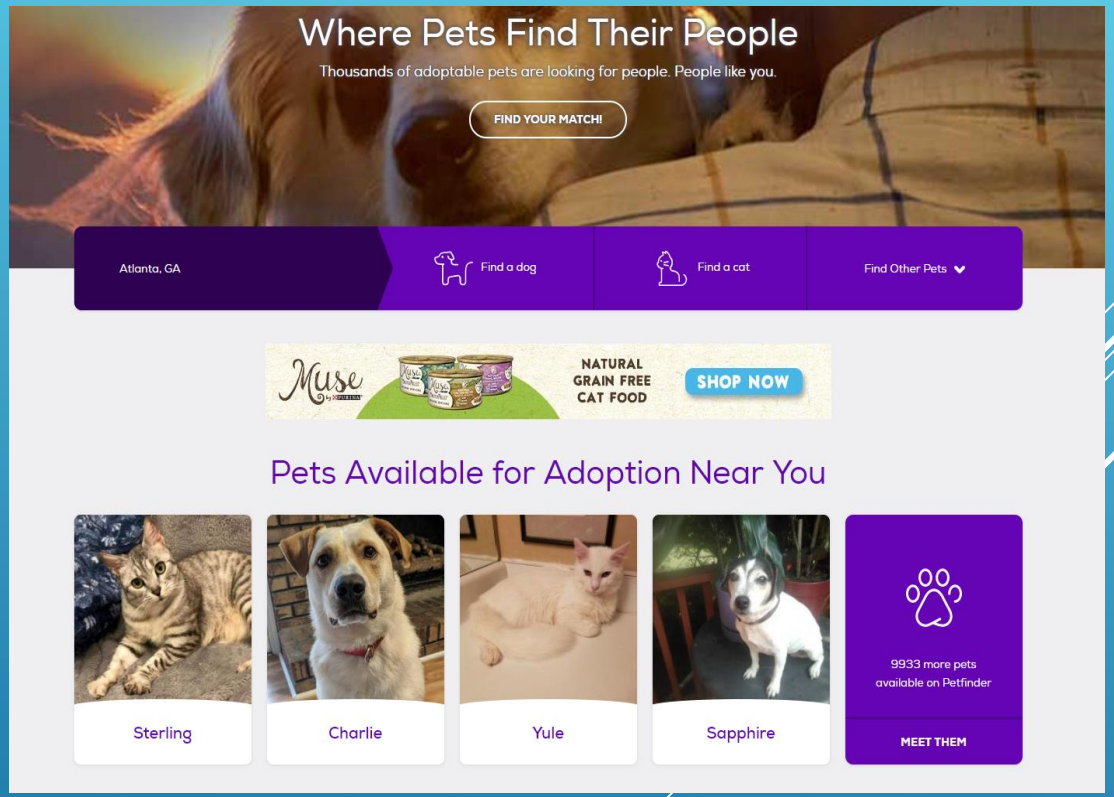
CAN AN ANIMAL'S ADOPTION BE PREDICTED FROM ITS PETFINDER PROFILE?

Julia Raykin

Final Capstone Presentation

THE PROBLEM

- Millions of stray animals are in shelters in dangers of being euthanized worldwide (World Health Organization)
- Petfinder – possible solution?
  - Brings together data on animals in local animal shelters
- *Research Question* – how do we improve Petfinder profiles and increase adoption rates in shelters?



DATA SOURCE

- Data from PetFinder.com will be analyzed to determine how an animal's PetFinder profile affects the rates at which animals get adopted.
- The dataset was obtained from Kaggle (https://www.kaggle.com/c/petfinder-adoption-prediction/data)

GOALS

- Determine which animals get adopted fastest
  - Identify important features in Petfinder profiles
  - Develop a model to predict an animal's adoptability
  - Tune the model features to improve its capabilities

AVAILABLE DATA

- Numerical features – age, quantity, fee, number of uploaded videos, number of uploaded photo, maturity size, fur length, health
- Categorical features – dewormed, vaccinated, sterilized, breed, state, gender, color
- Text – description of the animal
- Images of the pets

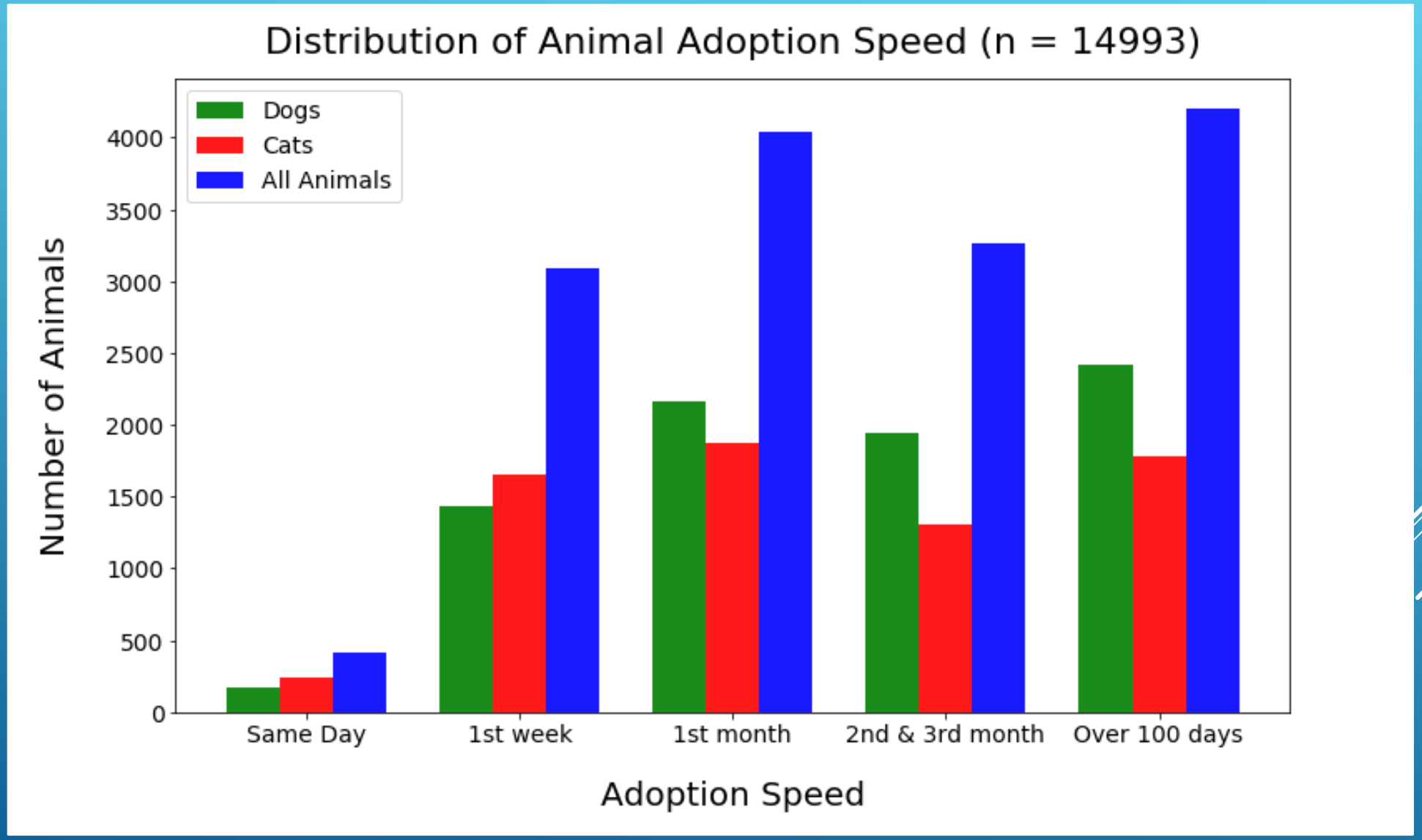
SAMPLE DATA

Type	Name	Age	Breed1	Breed2	Gender	Color1	Color2	Color3	MaturitySize	...	Health	Quantity	Fee	State
2	Nibble	3	299	0	1	1	7	0	1	...	1	1	100	41326

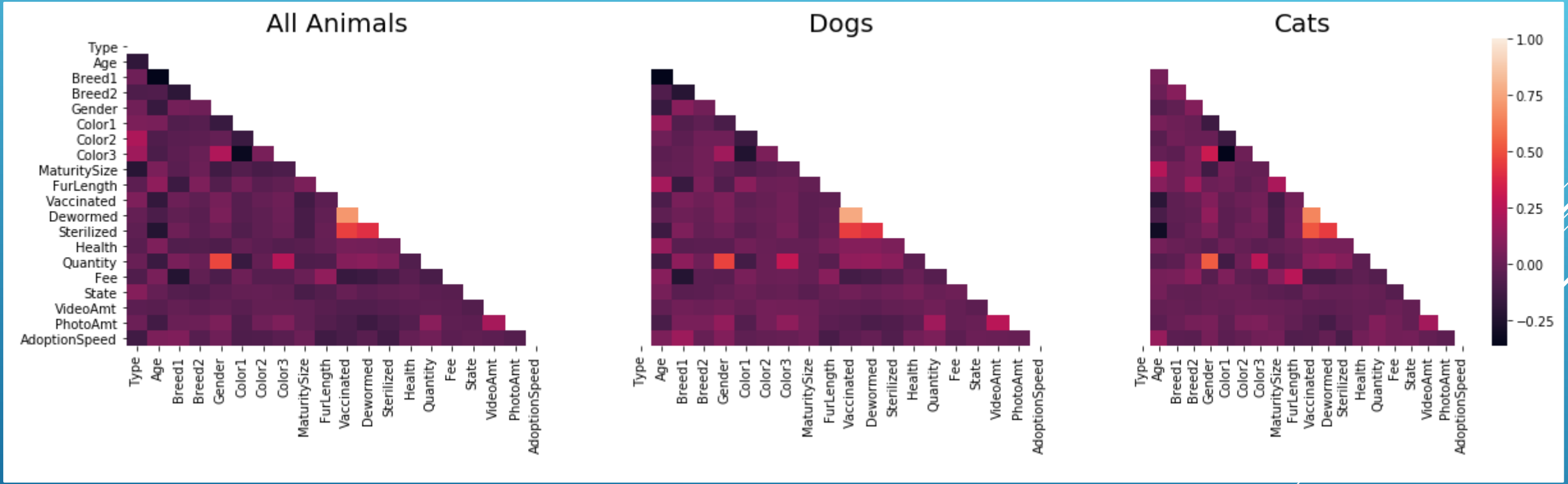
Quantity	Fee	State	RescuerID	VideoAmt	Description	PetID	PhotoAmt	AdoptionSpeed
1	100	41326	8480853f516546f6cf33aa88cd76c379	0	Nibble is a 3+ month old ball of cuteness. He ...	86e1089a3	1.0	2

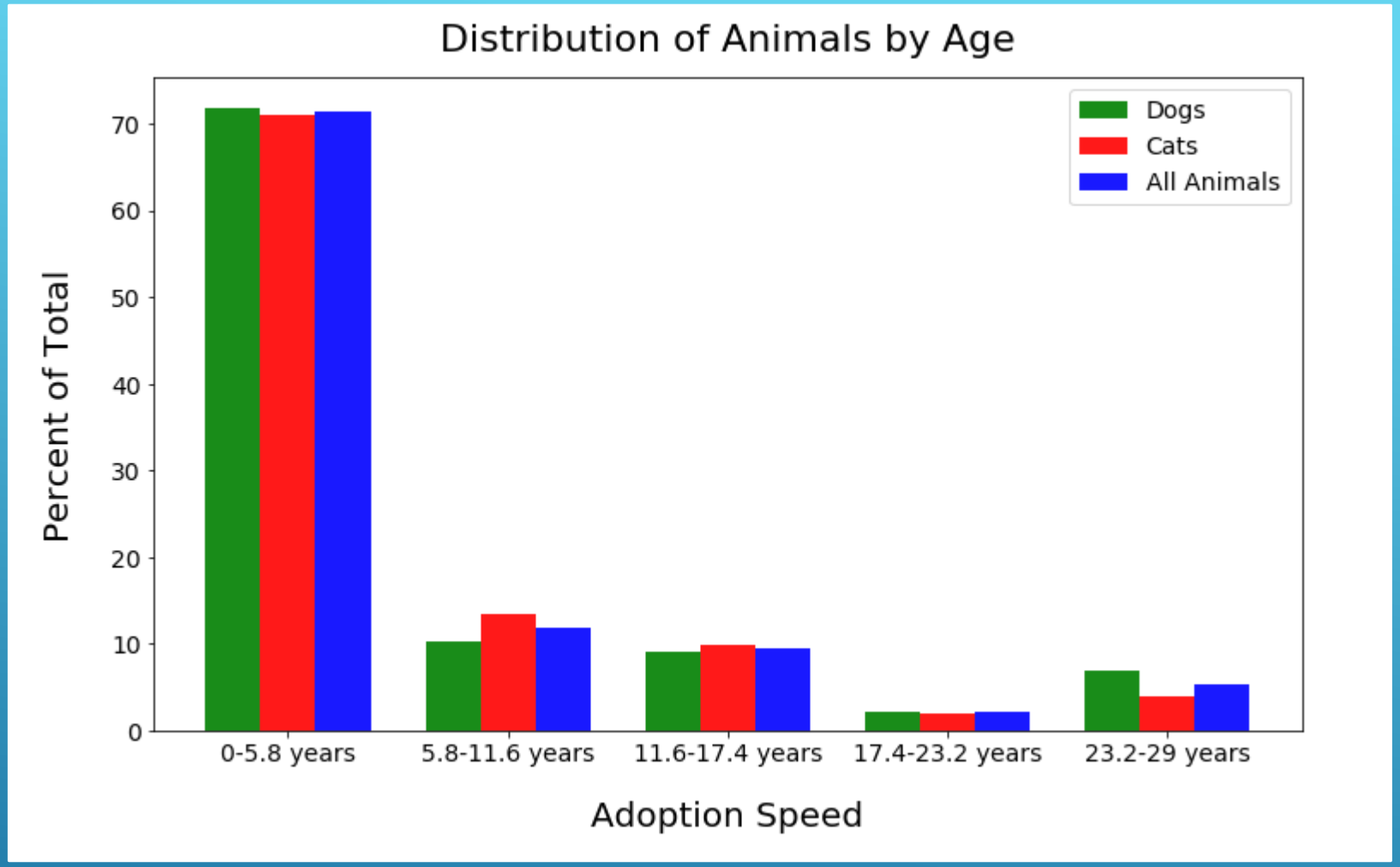
"Nibble is a 3+ month old ball of cuteness. He is energetic and playful. I rescued a couple of cats a few months ago but could not get them neutered in time as the clinic was fully scheduled. The result was this little kitty. I do not have enough space and funds to care for more cats in my household. Looking for responsible people to take over Nibble's care."

ADOPTION SPEED



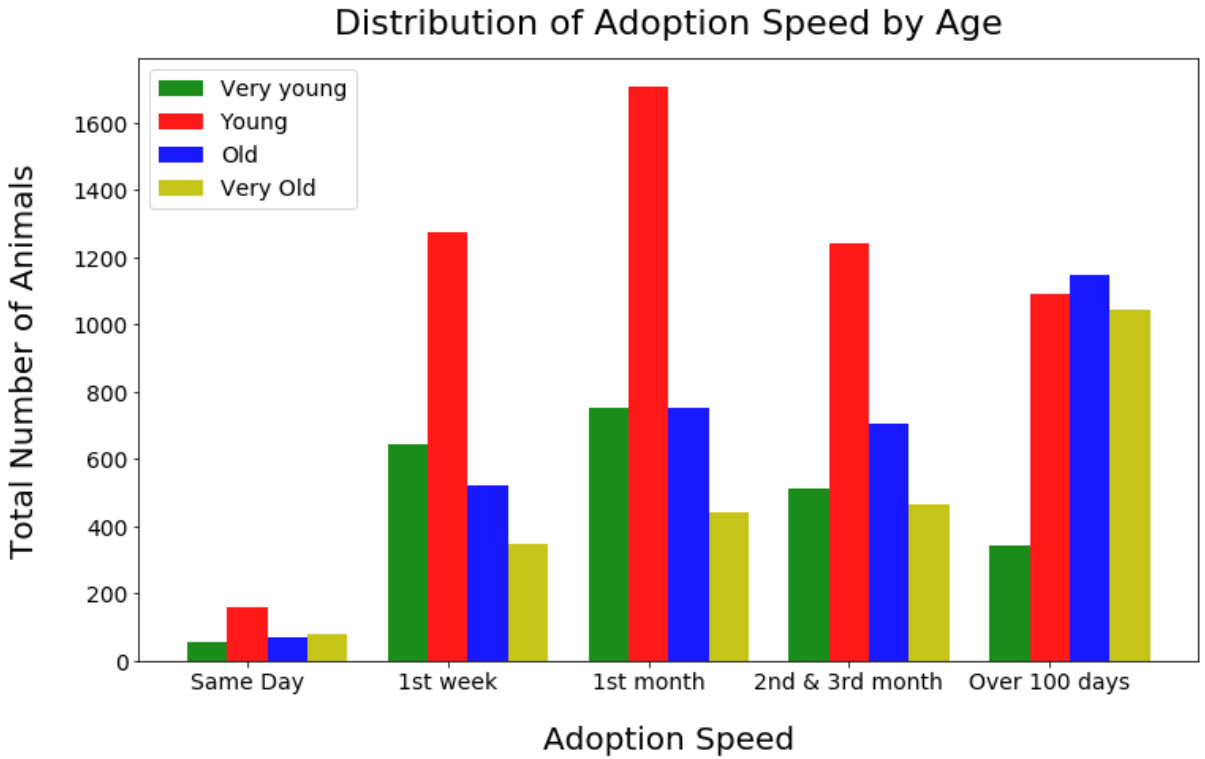
WHICH FACTORS ADOPTION SPEED?

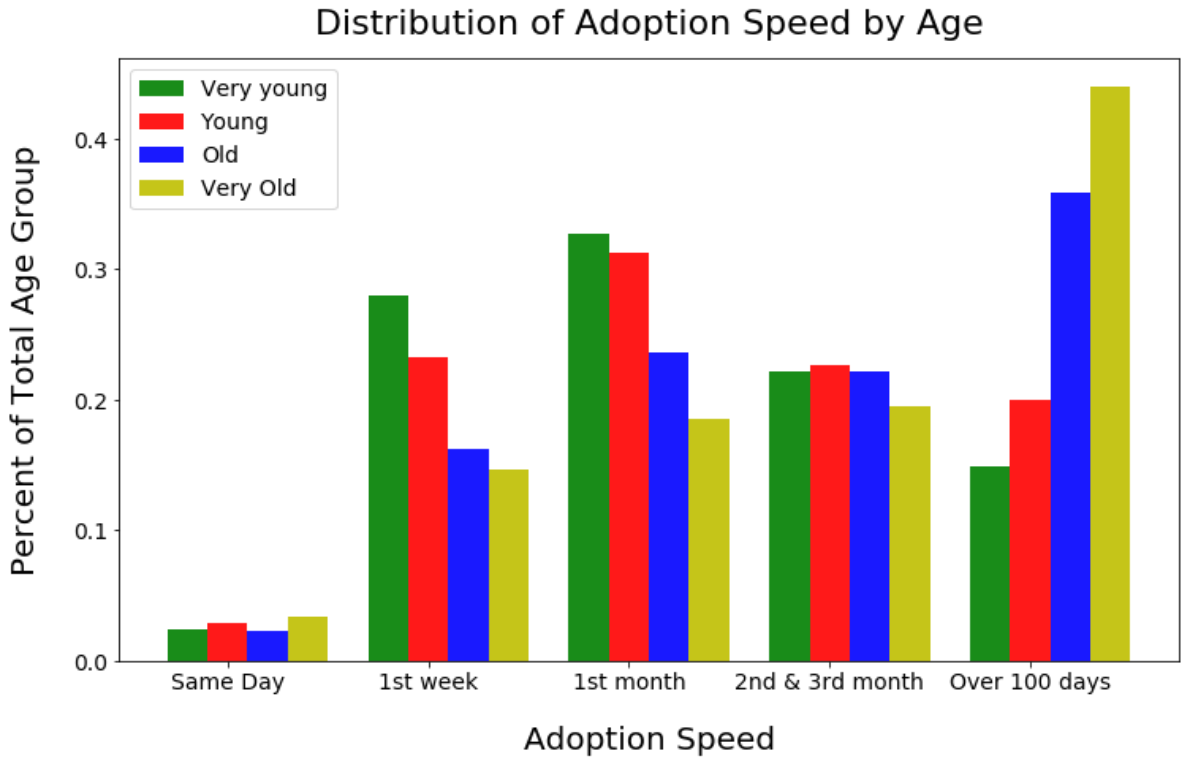




Most animals were under 5.8 years old. To get a better age distribution animals were binned by the following age ranges:
Very young: Under 1
Young: 2-4
Old: 4-10
Very old: older than 10

HOW DOES AGE AFFECT RATE OF ADOPTION?



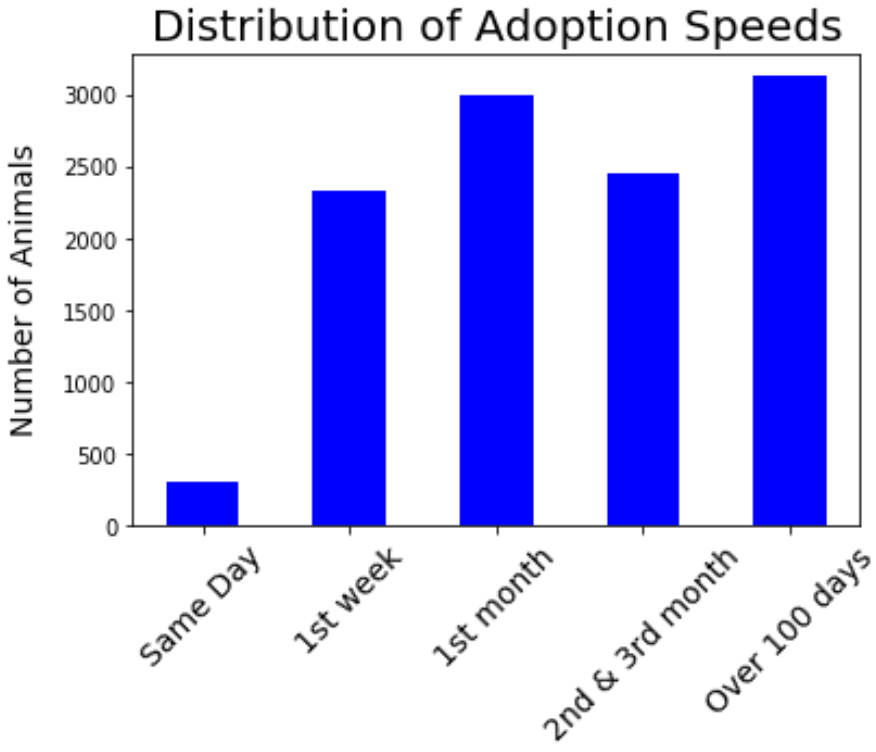


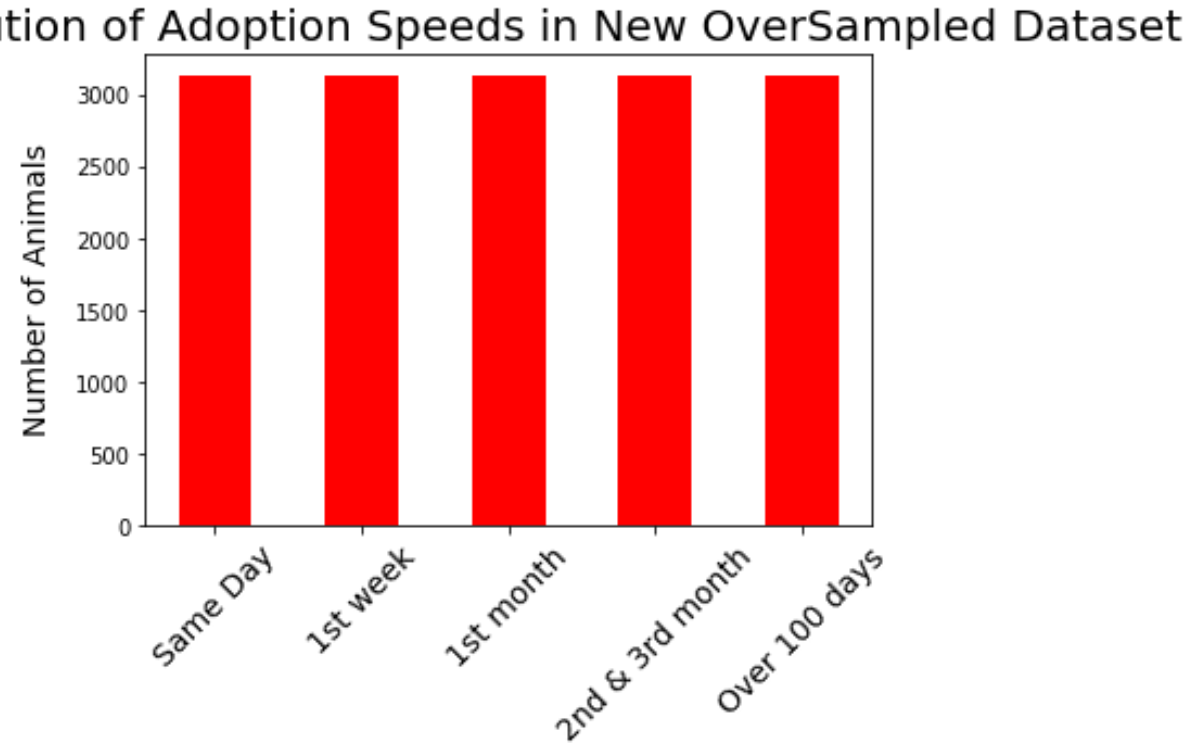
A higher percentage of older animals remain unadopted after 100 days.

MODEL 1

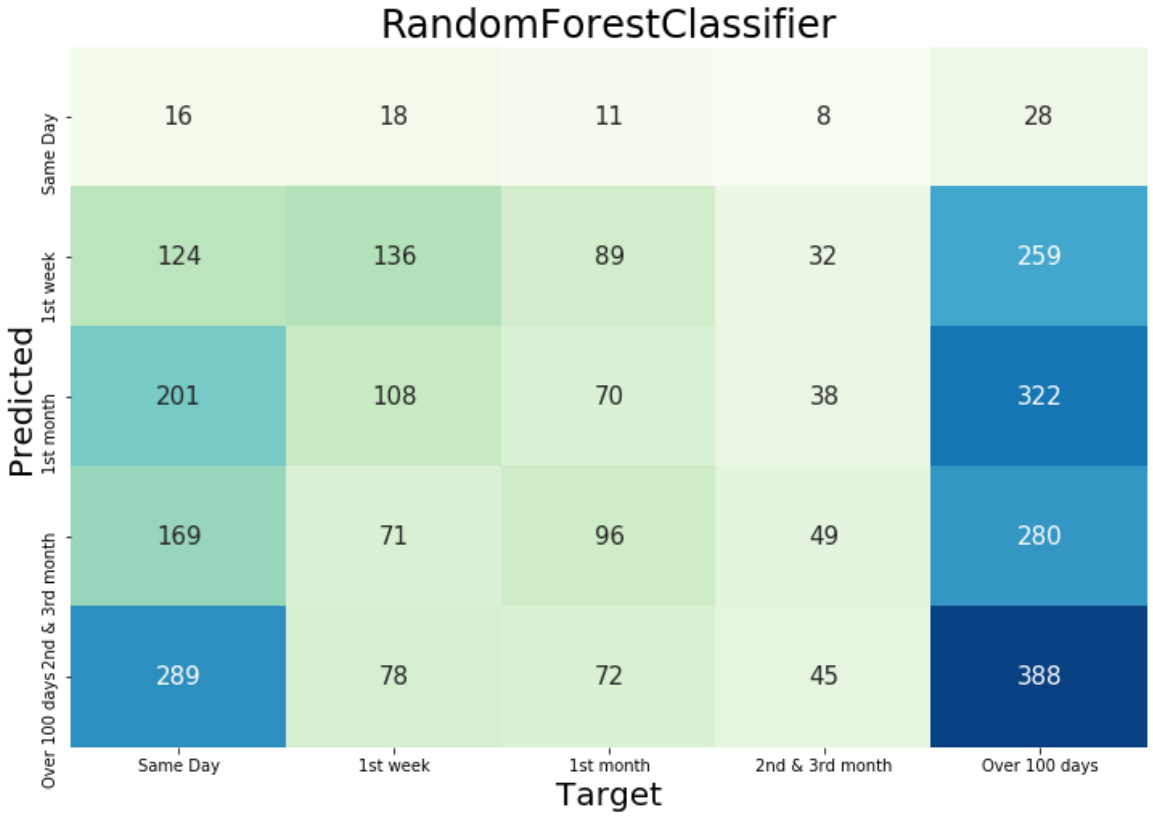
- One hot encode all categorical features, standardize all continuous features
- Split the data into training and testing datasets
- Try models:
  - Random Forest Classification
  - Xgboost Classification

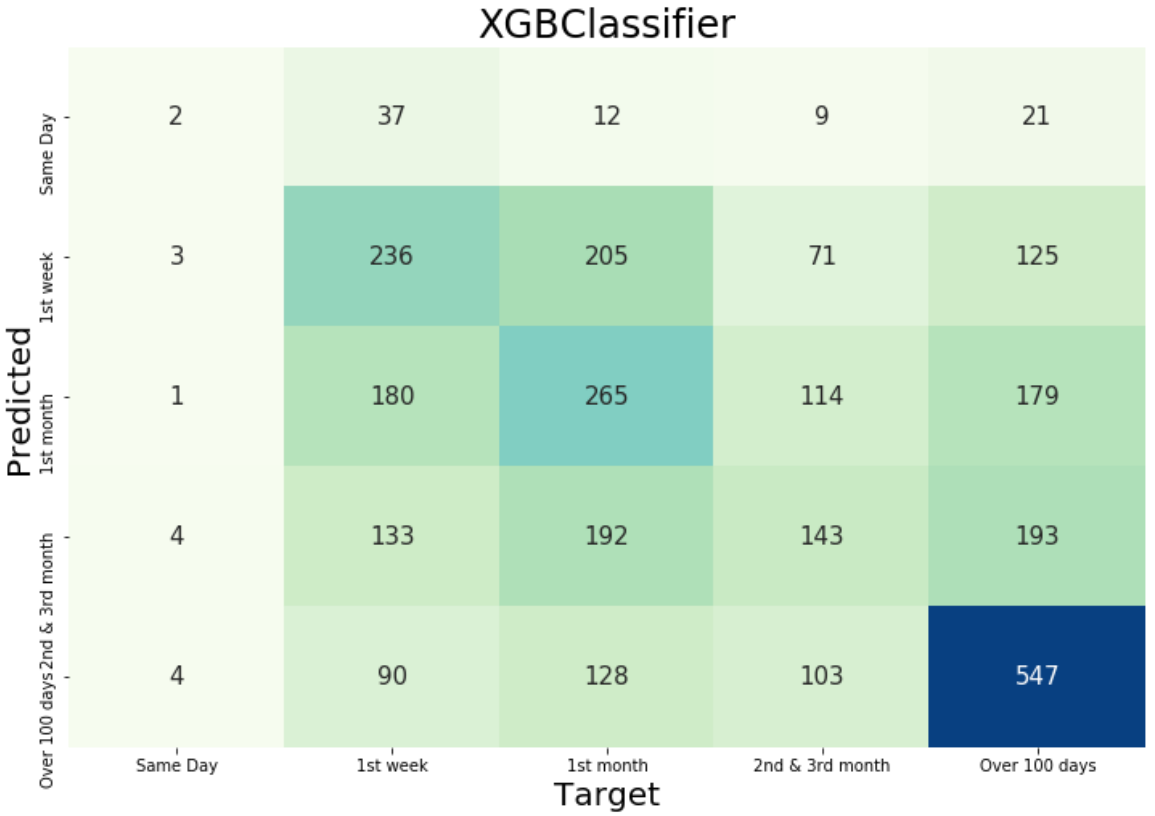
SMOTE WAS USED TO OVERSAMPLE THE TRAINING SET





MODEL 1 RESULTS



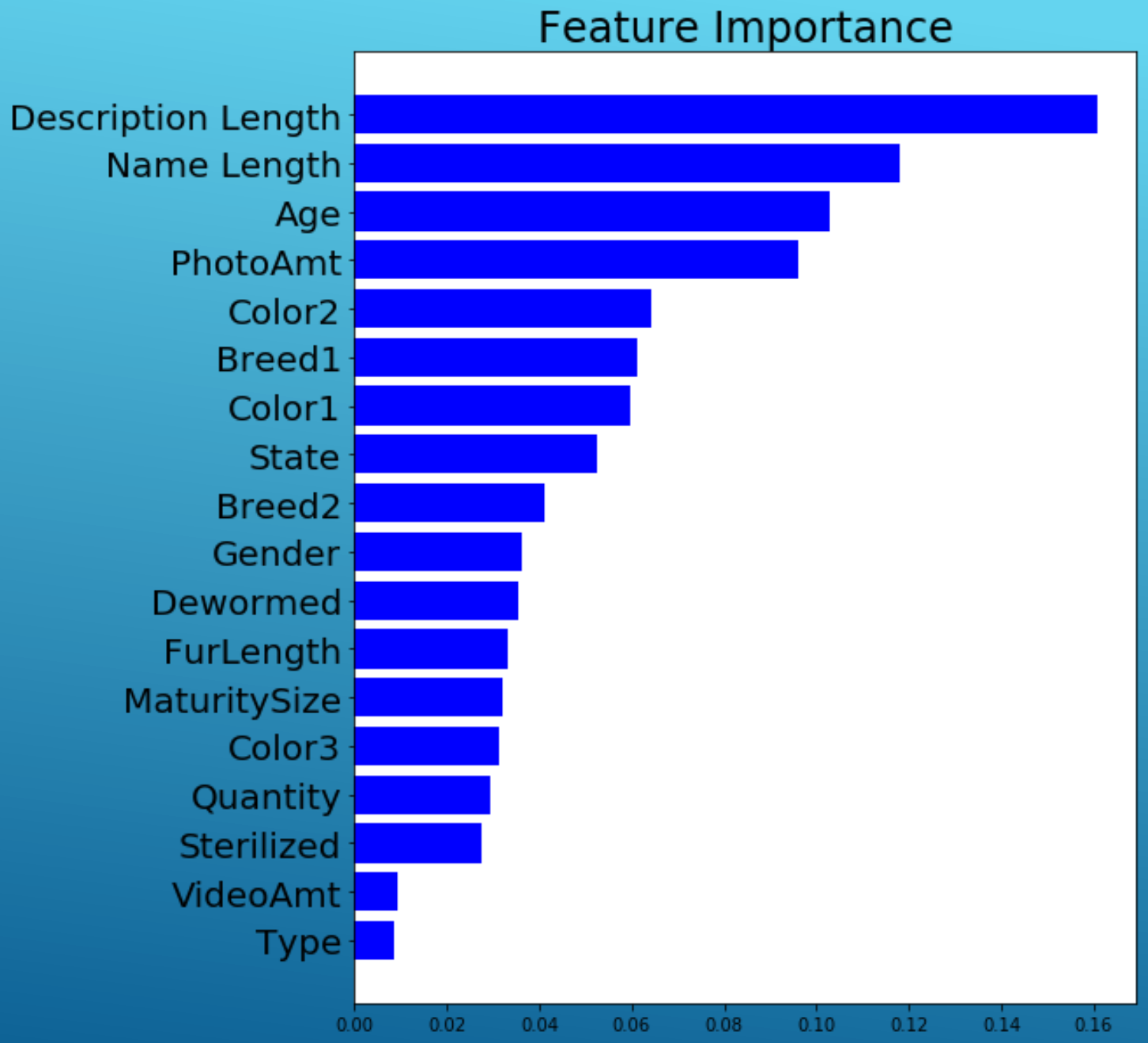


Training set accuracy: 0.99
Validation set accuracy: 0.36

	precision	recall	f1-score	support
0	0.17	0.17	0.17	81
1	0.32	0.40	0.36	640
2	0.31	0.33	0.32	739
3	0.28	0.25	0.27	665
4	0.54	0.46	0.50	872

Training set accuracy: 0.51
Validation set accuracy: 0.4

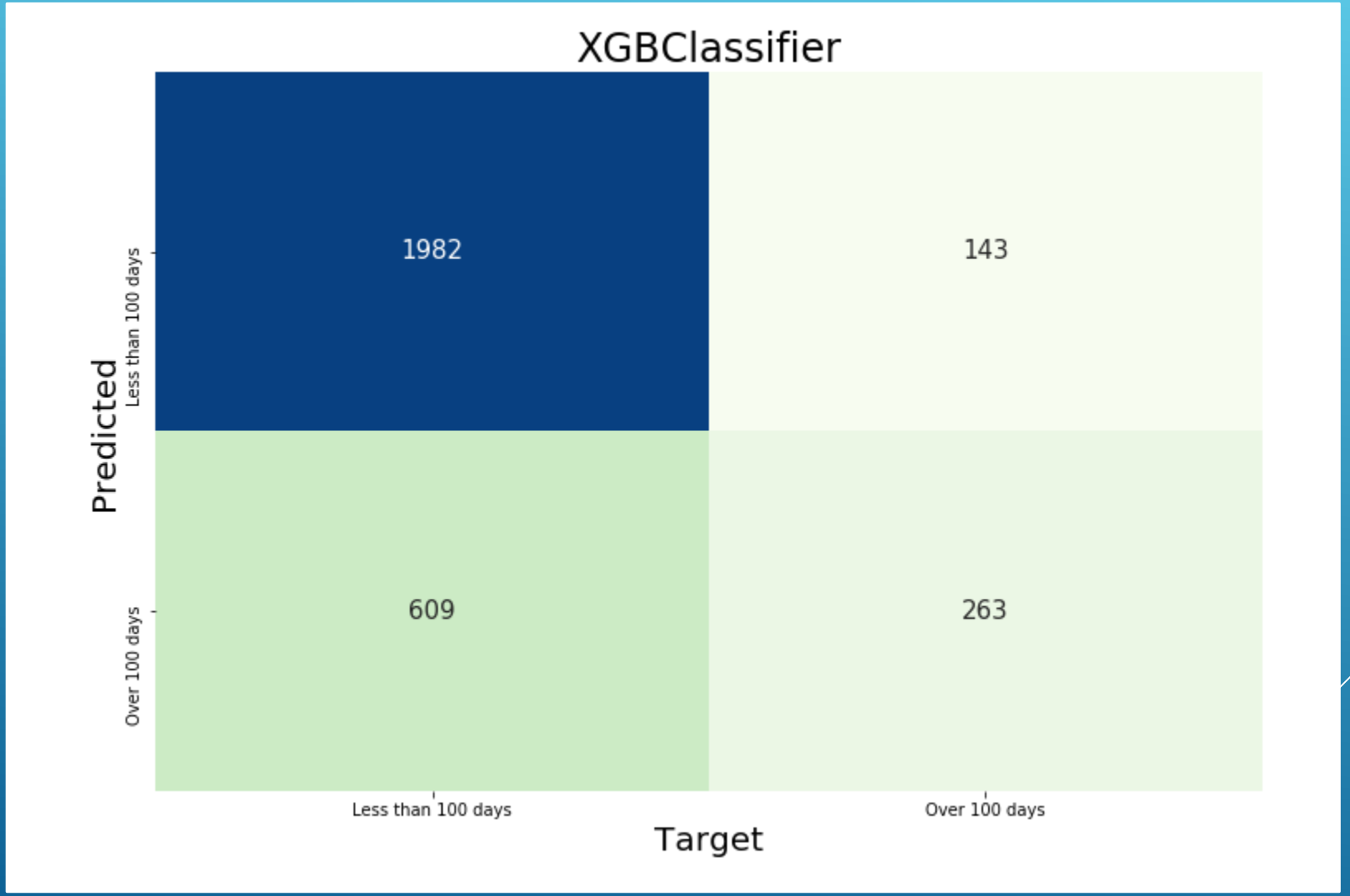
	precision	recall	f1-score	support
0	0.14	0.02	0.04	81
1	0.35	0.37	0.36	640
2	0.33	0.36	0.34	739
3	0.33	0.22	0.26	665
4	0.51	0.63	0.56	872


Feature Importance
Description Length
Name Length
Age
PhotoAmt
Color2
Breed1
Color1
State
Breed2
Gender
Dewormed
FurLength
MaturitySize
Color3
Quantity
Sterilized
VideoAmt
Type
0.00
0.02
0.04
0.06
0.08
0.10
0.12
0.14
0.16

MODEL 2

- Try Converting Adoption Rate to Binary
  - 0 if adopted faster than 1 week
  - 1 if longer than 1 week
- Xgboost

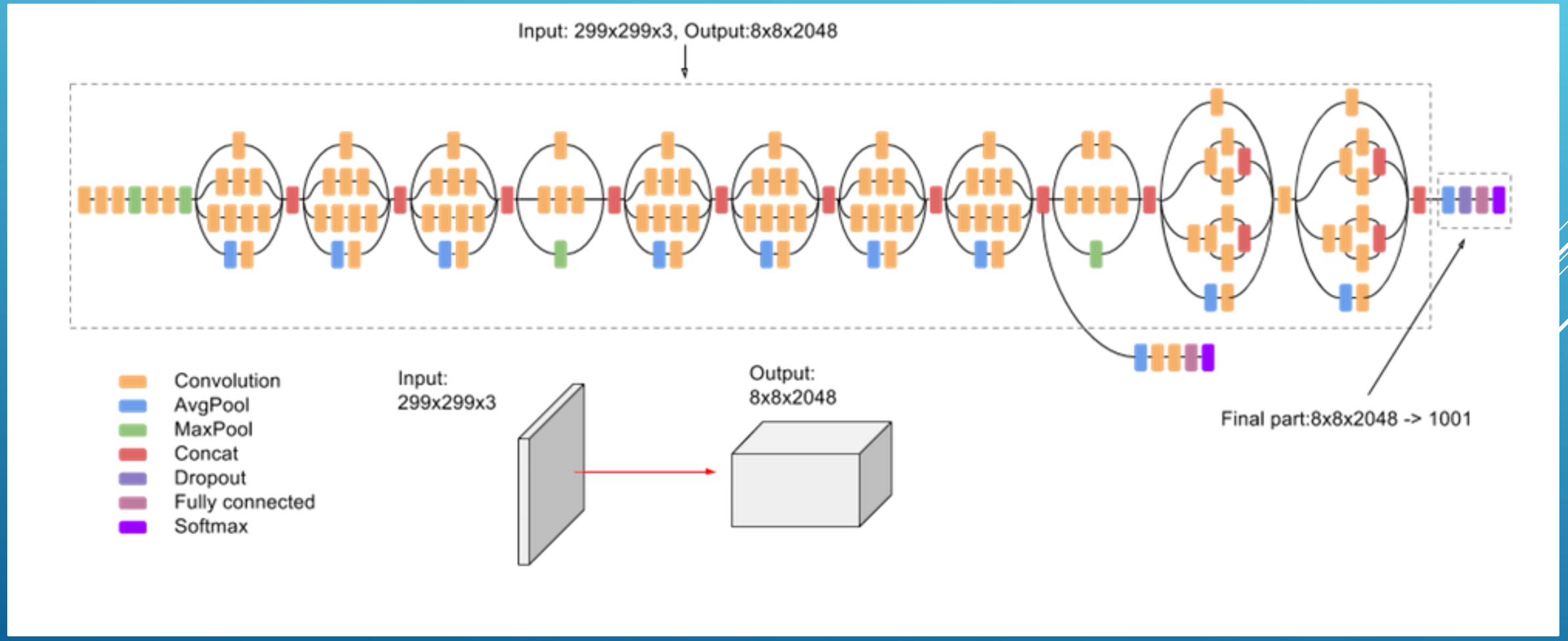
MODEL 2 RESULTS



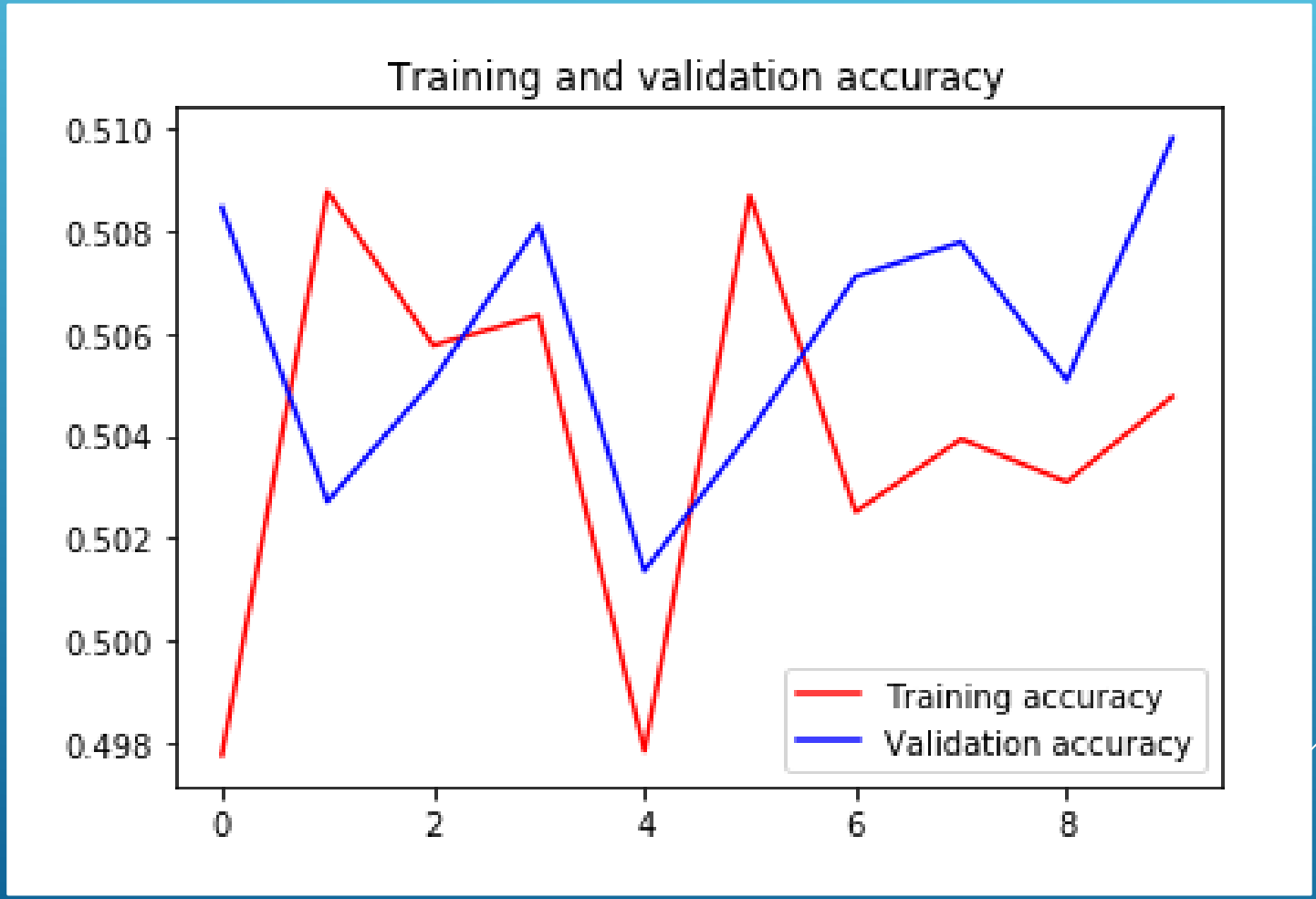
OTHER ATTEMPTED METHODS

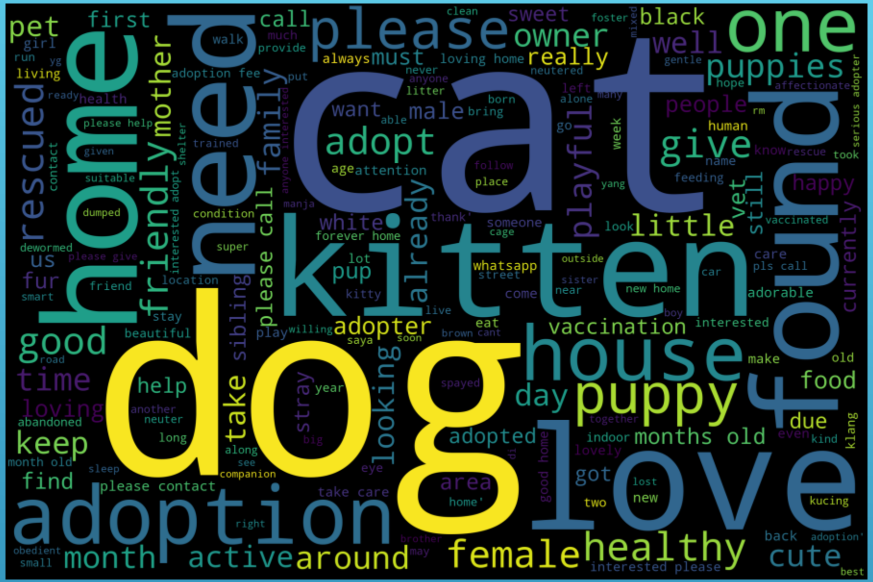
- SelectKbest features
- PCA
- Hyperparameter tuning

PRETRAINED CNN - INCEPTION

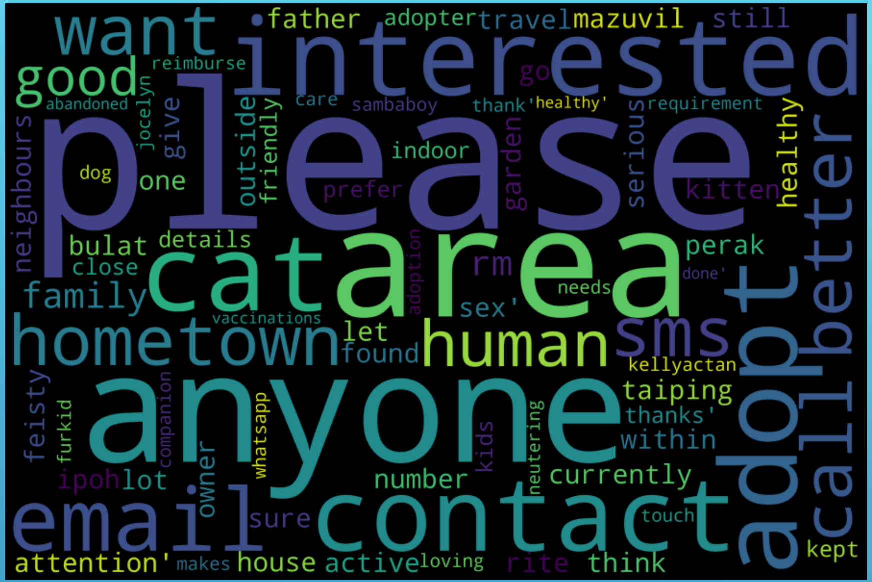


INCEPTION MODEL RESULTS

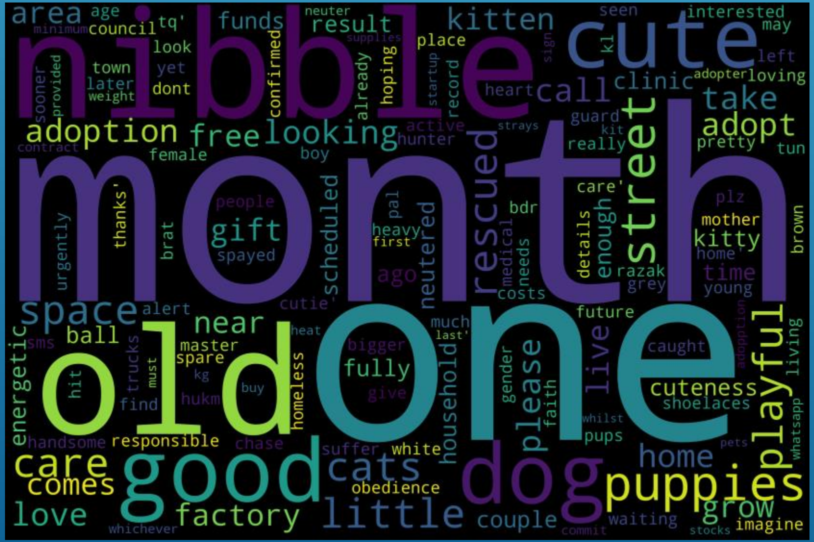




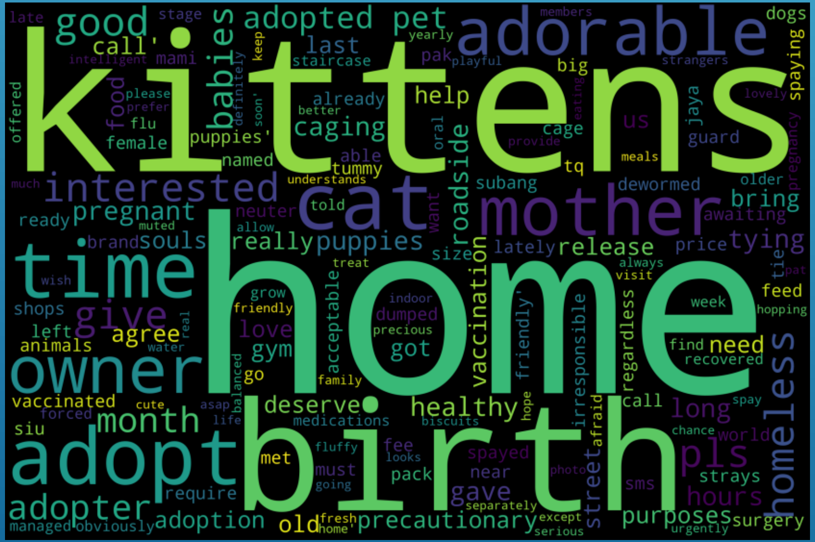
Same Day



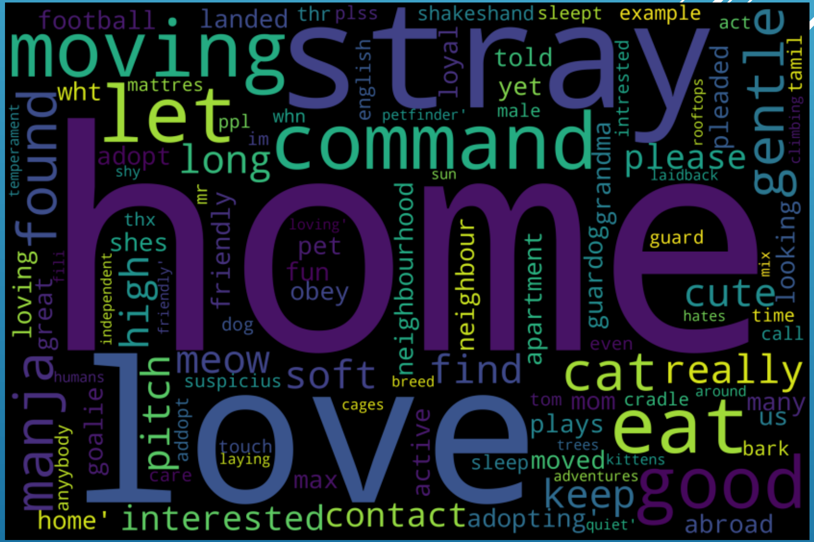
Less than 1 week



1st Month

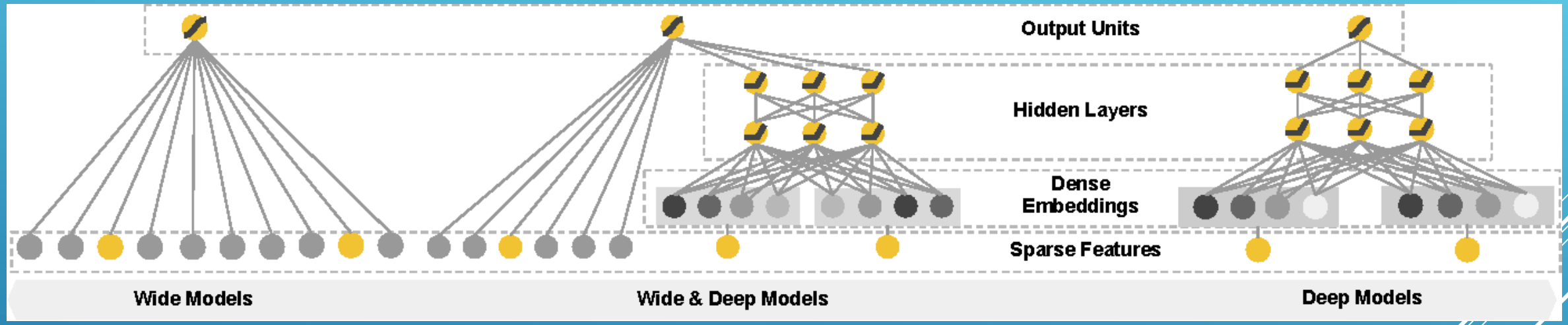


2nd and 3rd Month



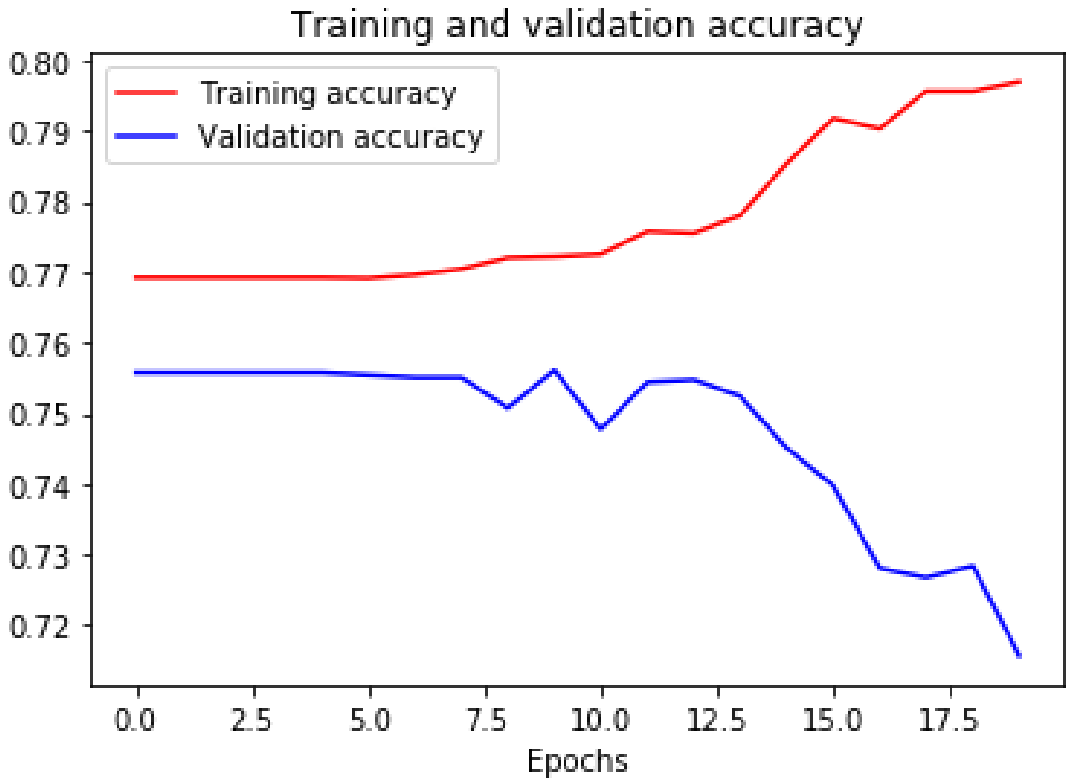
More than 100 days

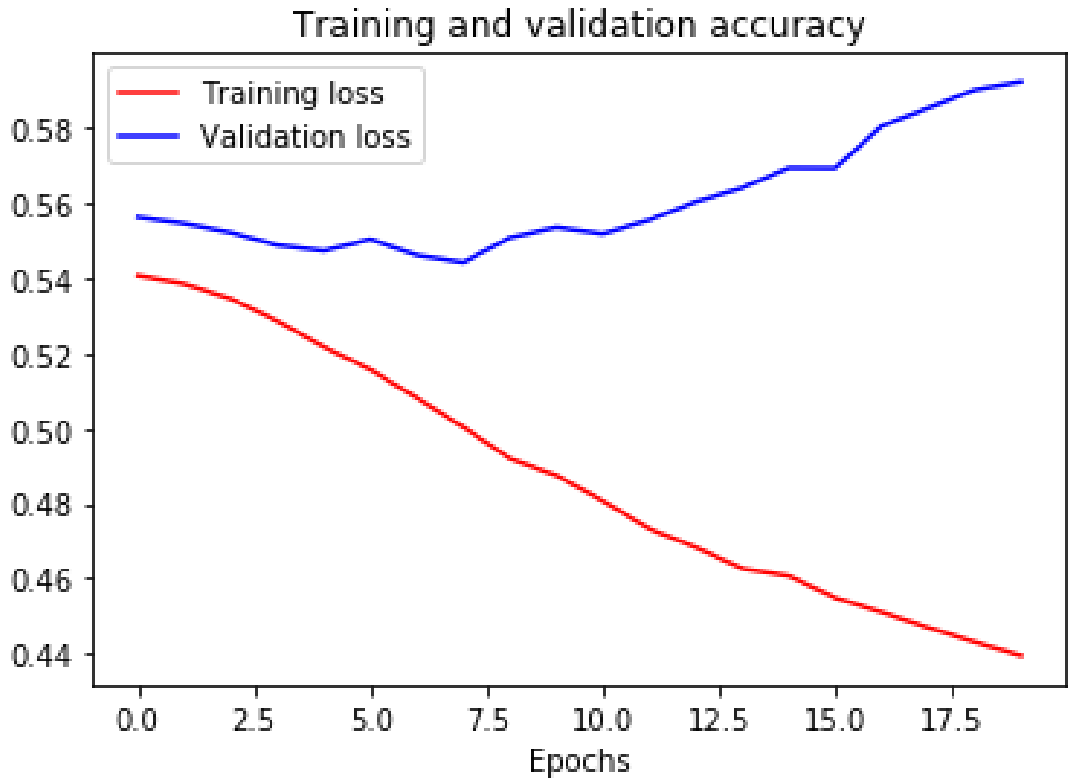
WIDE AND DEEP MODEL



Results:
Training set accuracy: 0.90
Validation set accuracy: 0.63

FINAL RESULTS





Test Accuracy: 78.5%

CONCLUSIONS

- The wide and deep model provided the best predictive accuracy
- Not much predictive information could be obtained from the pictures
- Number of photos provided is an important controllable feature
- Animal shelters should always post more than 2 pictures per animal

PROPOSED FURTHER RESEARCH

- Expand model to include data for other countries
  - This dataset just looks at adoptions in Malaysia
- Look at trends in U.S. by county and state
- Further analyze the descriptions to better define what makes a good prediction