

# PetFinder.my Adoption Speed Prediction Using Pet Profiles

Iris (Yuhui) Wang (iris1121@Stanford.edu), Eva (Yifan) Gong (gong717@Stanford.edu)  
 Advised by Gael Colas

## Problem Statement and Data

The goal of this project is to build a neural network to predict how fast a pet is adopted so that shelters/adoption agencies can better focus their resources to help the pet to find new homes. Our final model is a 2 layer neural network with 400+ input features. The overall accuracy of the 5 classes predictions is ~38% on the validation dataset. Test dataset scored a weighted kappa of 0.329.

The datasets are provided by Kaggle [1]. There are 14993 pets in the training set and 3948 pets in test set. For pets in training set, adoption speed are labeled as 5 different categories. Pet profiles include basic information like type, age, breed etc. as well as photos and text descriptions. Following field best practices, we randomly sampled 20% of the training set as the hold-out validation set.

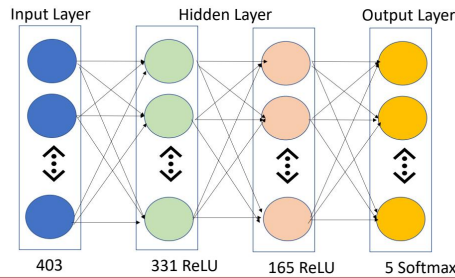


Class Distributions for Train and Validation Sets

## Features and Model

The raw input data provides 14 categorical variables and 5 continuous variables. We applied one-hot-encoding to the categorical variables. Since there are two variables with more than 100 categories, this significantly increased the number of features. We derived three variables from the text descriptions: 1) Document sentiment score 2) Document sentiment magnitude 3) Total number of letters in the text description. These features are derived because we think the pet's attractiveness is impacted by its description, especially the sentiment.

We built a neural network with 2 hidden layers. L2 regularization is used to control overfitting. Hyperparameters are tuned using random search. The final hyperparameters picked are: Learning rate=1.00E-05; Minibatch size=64; Regularization beta=0.04.



## Results and Future Work

For this softmax classifier, our main success metric is the overall prediction accuracy rate, defined as the total number of correctly predicted instances over the total number of instances. Following common multi-class performance measures outlined by Sokolova et al., we then calculated per-class and average Precision and Recall Rates through macro-averaging [2].

Adoption Speed	Overall Accuracy	Precision	Recall
Same Day	39%	0%	0%
1 to 7 Days		32%	44%
8 to 30 Days		31%	34%
30 to 90 Days		42%	14%
More than 90 Days		48%	64%

For future analysis, we plan to work on: 1) Building a Deeper network 2) Evaluating models based on non-accuracy metrics 3) Including image data as features

## Reference

- [1] PetFinder.my Adoption Prediction: How cute is that doggy in the shelter? Kaggle Competition. <https://www.kaggle.com/c/petfinder-adoption-prediction/>
- [2] Heaton, J. (2009). Introduction to neural networks with Java. C3esterfield (MO, USA): Heaton Research.
- [2] Sokolova, M., Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. Information Processing Management, 45(4), 427-437.