

Predicting Adoption Rates of Shelter Animals

BY MAX PIASEVOLI AND CAMDEN OLSON

Introduction

6.5 million animals end up in U.S. shelters every year

Almost a 4th of them will end up being euthanized due to overcrowding in shelters

This dilemma is pervasive throughout the world, not just in the U.S.

Many parties are interested in what factors improve an animal's chances of adoption.

Goal:

1. PREDICT ADOPTION SPEEDS

2. IDENTIFY FACTORS LEADING TO QUICKER ADOPTION

Related Work:

ADOPTION VS EUTHANASIA STUDY

- USED MULTIPLE LOGISTIC REGRESSION TO PREDICT LIKELIHOOD OF ADOPTION
- AGE, SEX AND COAT COLOR ARE CORRELATED WITH ADOPTION

LENGTH OF STAY (LOS) STUDY

- OBSERVED PHENOTYPIC CHARACTERISTICS IN RELATION TO ADOPTION SPEED
- LOS INCREASES LINEARLY WITH AGE
- SEX AND COAT COLOR ARE UNCORRELATED WITH LOS

Data



11,994 Petfinder profiles of adoptable cats and dogs in Malaysia



Petfinder = online animal adoption cite



Each profile includes information on age, breed, gender, color, maturity, size, fur length, health conditions, state in Malaysia and a description of the pet



Predict adoption speed, broken into 5 categories

0 = same day, 1 = within 1st week, 2 = 1st month, 3 = 2nd and 3rd month, 4 = no adoption after 100 days

Classifiers and Performance Metrics

Classifiers

1. Random Forest Classifier (RF)
2. Decision Tree Classifier (DT)
3. Logistic Regression with L1-Norm Classifier (LR1)
4. Logistic Regression with L2-Norm Classifier (LR2)

Performance Metrics

1. Accuracy
2. Precision
3. Recall
4. F1-Score

Categorical Data Classifiers

Classifier	RF	DT	LR2	LR1
Accuracy	0.368	0.344	0.390	0.380
Precision	0.377	0.339	0.596	0.596
Recall	0.368	0.344	0.390	0.380
F1-Score	0.372	0.341	0.459	0.454

- Logistic Regression with L2-Norm performed best across all metrics
- Age and # of pets on the listing were predictive for adoption speed 4
- Poodles, which are the most popular dog breed in Malaysia, were predictive of adoption speed 0
- Photo amount was important for predicting adoption speeds 2 and 3

Misclassification Analysis

Methods

- Predicted adoption speed using LassoCV linear regression to calculate residuals
 - Selected best features that predicted the residuals
 - Allowed us to identify features that led to over and under predictions of adoption speed
-

Analysis

- Top 5 features most responsible for residuals:
 - Dog Breed: Mixed Breed; Maturity Size: Medium; Not Vaccinated; Sterilized: Yes; Sterilized: No
- Being mixed breed led to a higher likelihood of being under predicted. Not being a mixed breed led to a higher likelihood of being over predicted.
- Medium-sized dogs were under predicted.
- Sterilization and Vaccination status did not show significant trends but did throw off predictions.

Text Data Classifiers

Classifier	RF	DT	LR2	LR1
Accuracy	0.365	0.340	0.371	0.363
Precision	0.38	0.343	0.384	0.376
Recall	0.365	0.340	0.371	0.363
F1-Score	0.371	0.341	0.376	0.368

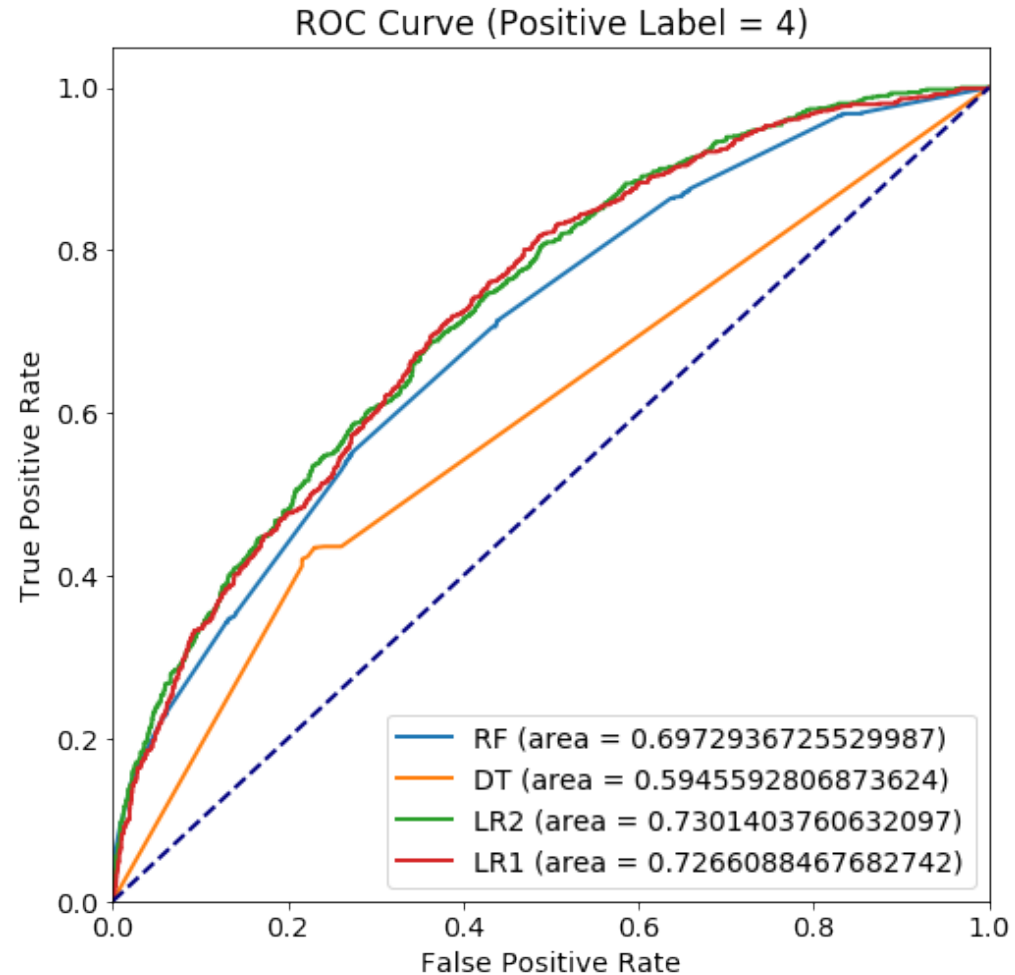
- Logistic Regression with L2-Norm performed best across all metrics
- Words indicating illness or injury were typically correlated with slower adoption speed
- Positive words generally accompanied each of the class labels

Combined Data Classifiers

Classifier	RF	DT	LR2	LR1
Accuracy	0.418	0.360	0.371	0.369
Precision	0.444	0.359	0.580	0.610
Recall	0.418	0.360	0.371	0.369
F1-Score	0.428	0.359	0.443	0.451

- Random Forest performed best across all metrics
- Logistic Regression classifiers have the best precision scores
 - Only predicted labels of 2, 3, and 4
 - Precision metric is skewed due to no prediction of 1 or 2 labels

ROC Curves for Combined Classifier



- Good ROC Curves: 0, 1, 4
- Bad ROC Curves: 2, 3

Classifier	RF	DT	LR2	LR1
0 Label AUC	0.681	0.578	0.518	0.508
1 Label AUC	0.666	0.564	0.581	0.579
2 Label AUC	0.606	0.549	0.596	0.587
3 Label AUC	0.617	0.554	0.595	0.591
4 Label AUC	0.763	0.634	0.680	0.675

Discussion

- Logistic Regression with an L2 norm was the best at predicting adoption speed with categorical only or text only data
- Random forest was best at predicting adoption speed with categorical and text data combined
- Age and illness was correlated with slower adoption speed

Future Work

- Feature engineering to account for purebred status, level of training, etc.
- Analyze image data
- Look at Petfinder data sets in other countries and see if trends are similar