Predicting Animal Adoptability from Key Indicators Abstract

Jessica Buzzelli, Ethan Scott, Trent Davidson, Brian Davis

Project Description

Our group wants to look at animal shelter data from Austin Animal Center and create a tool to predict outcomes. Days until an animal is adopted is a largely uncontrollable process since the shelter has no say in the "quality" of animals that they receive, so we aim to apply the outcome prediction tool to determine a "pricing methodology" for adoption fees such that animals with a shorter predicted turnover time have higher fees. By moving away from a single adoption fee for all animals, the animal shelter will be able to keep its mean adoption revenue over a given time period more in control.

Methodologies

With knowledge of how long it takes an animal to be adopted and the traits of the animals, we will conduct a regression analysis (comparing linear and logistic methods) to determine a "formula" to apply to new data to estimate the number of days until they are expected to be adopted given their physical characteristics. After building an initial model, we will determine the relative significance of each trait in the adoption selection process and refine our model by eliminating the least significant traits or condensing the number of traits with a method such as singular value decomposition or principal component analysis. After building a prediction model, we will create a "pricing rule" based off of our final regression formula to determine the fee of each new animal. We will evaluate the success of the proposed pricing rule based off of a control chart of the shelter's adoption revenue over time.

Data Set:

The data set contains 14993 instances each with 23 different features to identify information on the animals. The 10 features of interest for process analysis are: Type (dog or cat), Age (in months), Breed1+Breed2 (primary+secondary), Gender(Male,Female, or mixed if profile represents group of animals), MaturitySize(small, medium, large, extra large, not specified), Vaccinated(yes, no, or unsure), Dewormed(yes, no, or unsure), Steralized(yes,no, or unsure), Health(healthy, minor injury, serious injury, or not specified), Fee(amount in $ with 0=free).

Intended Solution

Using a regression-based model to predict days until an animal is adopted and an adoption fee structure where higher demand animals cost more, we aim to keep the expected adoption revenue within control limits.