Predicting the Outcome of Dogs at the Austin Animal Center

Background

The Austin Animal Center is the largest no-kill animal shelter in the United States. The dataset we worked on contains information about intakes and outcomes of animals entering the Austin Animal Center. We wanted to investigate variables such as breed, size, and sex to predict the outcome of dogs. Our results will inform shelters' decisions upon intake of a new animal, including how long it is expected to remain in the shelter, how likely it will be adopted, and deliberate choices to help increase the likelihood of adoption.

Methods

In order to repurpose the data for a regression tree, we calculated age upon outcome by computing the difference between an animal's date of birth and date of outcome. In addition, to reduce the number of categories within breed, we merged the current dataset with an additional one that sorts dog breeds into different sizes (toy, small, medium, large, and giant). Also, in order to facilitate and simplify prediction, we categorized all values in outcome type into four categories. Lastly, we coded two new variables—sex and whether the animal was spayed or neutered—based on information available in the original dataset.

We divided data into training and test subsets, each with 35,000 and 10,364 units. Then, we created regression trees predicting outcome type based on predictors such as age upon outcome, size based on breed, sex, whether the animals were spayed or neutered, condition upon intake, and the time they spent in the shelter.

Results

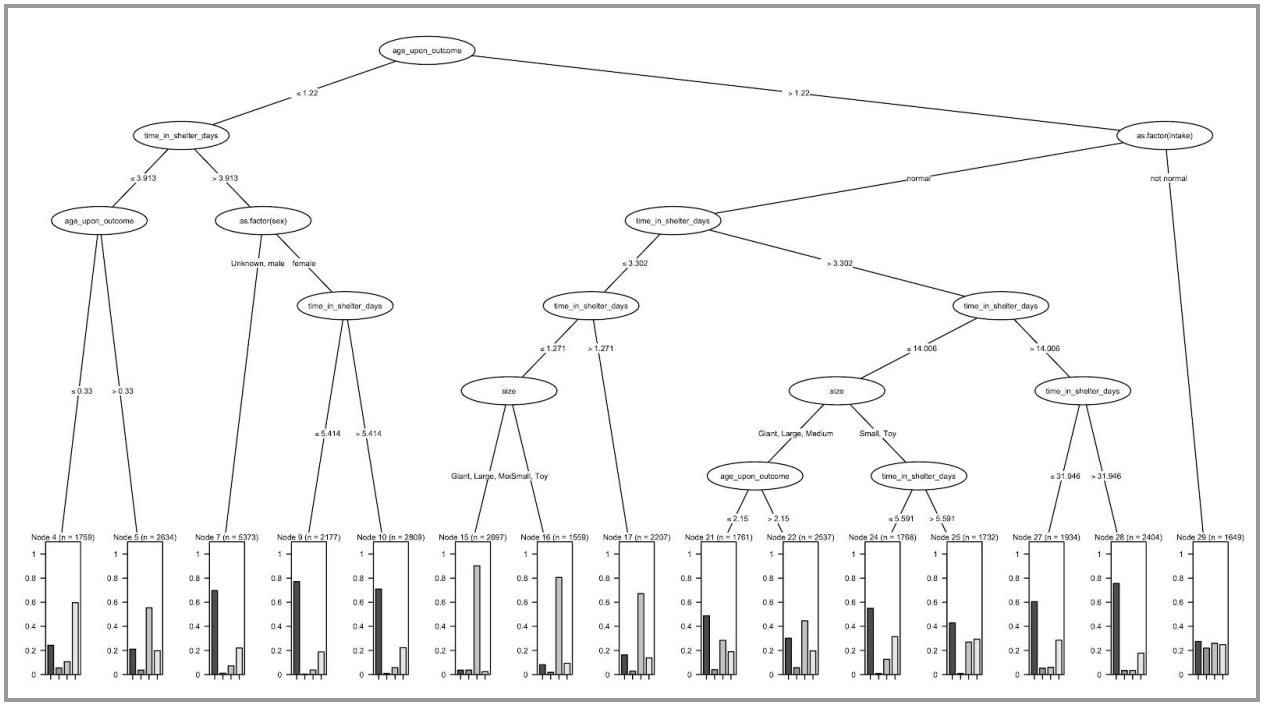

Figure 1: Regression tree based on the following predictors: intake condition, whether the animal is spayed or neutered, time in shelter, size, sex, and age upon outcome. The bar graphs respectively show the proportion of likelihood for whether the dog was adopted, died, returned, or transferred.

Based on the regression tree, age upon outcome is the most important predictor in predicting outcome, consistent with the intuition that younger animals tend to be adopted more easily. Within dogs that have a younger age upon outcome, those that spent longer in the shelter (i.e. were younger upon intake) had a higher likelihood of adoption. For dogs that were older upon outcome, intake condition has a high predictive power. When income condition is normal, adoption preferences seem consistent. Dogs that were younger upon intake and smaller size have a higher likelihood of getting adopted.

Sample size: 10364	Observed				
Validity: 66.9%					
Predicted		Adopted	Died	Returned	Transfer	**Total**
	Adopted	**4023**	114	701	1376	**6214**
	Died	4	**72**	51	23	**150**
	Returned	465	113	**2310**	362	**3250**
	Transferred	108	58	111	**472**	**749**
	Total	**4600**	**357**	**3173**	**2233**	**10363**

Figure 2: Validation table for **full model** using the following predictors: intake condition, whether the animal is spayed or neutered, time in shelter, size, sex, and age upon outcome.

Sample size: 10364	Observed				
Validity: 44.5%					
Predicted		Adopted	Died	Returned	Transfer	**Total**
	Adopted	**4387**	239	3028	2006	**9660**
	Died	4	**0**	0	0	**4**
	Returned	0	0	**0**	12	**12**
	Transferred	213	118	145	**227**	**703**
	Total	**4604**	**357**	**3173**	**2245**	**10379**

Figure 3: Validation table for **reduced model** using the following predictors: intake condition, size, and sex.

Discussion

Our research question was to find the best model for predicting the outcome type of animals. Above is the prediction tree using the full model. We set the minimum sum of units in a terminal node to 1,500 only for a clearer graphic representation. There are no such restrictions posed on the prediction and validation tables.

The full model correctly predicts 66.9% of the outcomes. The reduced model correctly predicts 44.5% of the outcomes. Our full model predicts at an accuracy level that is significantly higher than the reduced model. As a result, we ruled out the concern of overfitting. There is an observed tendency for the reduced model to underpredict "died" and "returned" outcomes due to the absence of age in the model. The middle age of those two categories to be higher than "adopted" and "transferred." From the prediction tree we can see that in most cases, there is a higher percentage of predict adoptions, as well as observed adoptions. There are less predicted adoptions when intake condition is not normal and sex is unknown. The predictors that have the largest influence in the tree are intake condition, size, and sex.

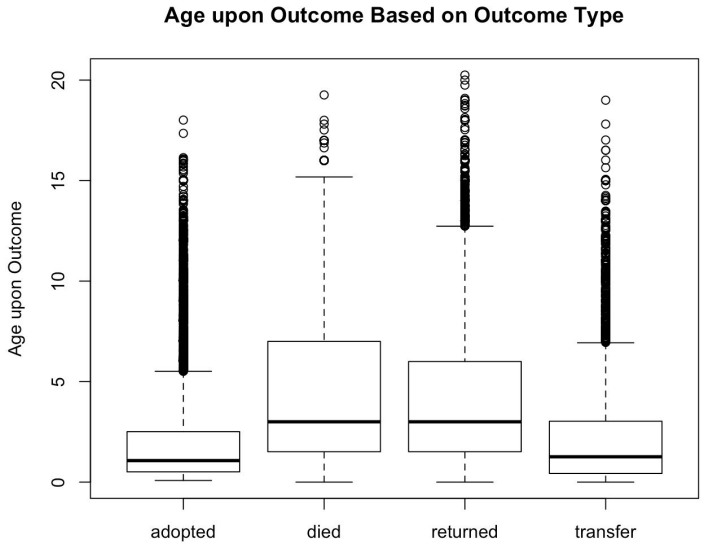

Figure 4: Age upon Outcome Based on Outcome Type

Conclusion

In the future, our model could be used to investigate other outcome variables such as time spent in shelter, outcome age, and days with the most activity. These results can be used by the shelter to determine which animals may have a hard time getting adopted and adjust their decisions. Also, predicting the animal's outcome at time of intake may save the shelter time and resources. However, there may be ethical issues when it is predicted that an animal is likely to be put down, as this may decrease a shelter's motivation to try their best to tend to its needs.

References

Our data source:
https://www.kaggle.com/aaronschlegel/austin-animal-center-shelter-intakes-and-outcomes