**Final Project**

Akorede Oyinlola, Pranav Khanna, Peilin Xin,

Harsh Atul Samani, and Samuel Klamar

Northeastern University

ALY 6070 - Communication and Visualization for Data Analytics

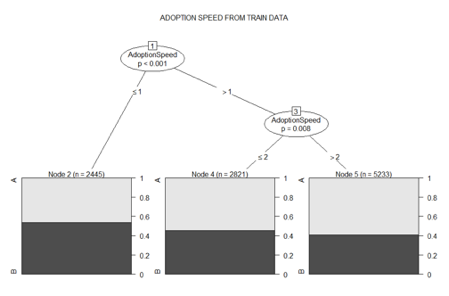
Instructor Name – Prof**.** Mykhaylo Trubskyy

Date- 31/3/2019



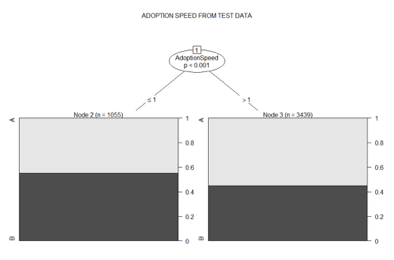
Adoption speed is a major limiting factor for pet shelters because faster adoptions lead to higher levels of revenue and lower pet housing costs. From an animal well-being perspective faster adoption are also beneficial as they shelter to take in more animals over the course of a year ("Resources – Koret Shelter Medicine Program", 2017). For these reasons’ shelters are always looking into ways better understand what gets pets adopted and especially what factors they can influence to help their own animals get adopted. In this project, several statistical models will be applied to the available data set to help us understand what factors the organization can focus on to help increase the adoption speed of pets within the facility. The variety of analytic techniques will allow for a broader understanding of the relationships between all factors.

In this project, we are performing classification model (Conditional Inference Tree) to predict the pet adoption speed. This model is chosen due to its flexibility, strong statistical foundation, as well as great capabilities to generalize and cope with problems in the data. To predict the adoption speed of the pet in days, letter A have been assigned to Dog and B as Cat. Our dataset is divided into train and test data at a ratio 70:30 (train data = 70% while test data=30%). This is done to help ensure both data sets are similar and reduce the effect of data discrepancies and better understand the characteristics of the model. Video amount and health are the variables chosen to test this classification based on the assumption that the pet are adopted based on the number of video which the agencies upload as ad to get the attention of those coming to adopt the pet. In addition, after making up your mind based on the videos seen, one other factor most people consider is how healthy the pet is. Unhealthy pet could result in both emotional cost due to death and financial cost (cost incurred from regular hospital visit). Below shows result from the test data:



The above shows that when adoption speed less than or equal to 1 (week 1: 0 to 7 days), there is a 55% chance of cat being adopted while dog has 45%.When the adoption speed is less than or equal to 2 (1st month), there is 45% chance of Cats being adopted while Dog is 55% and when the adoption speed is greater than 2 (more than one month), Dogs has 60% chance of adoption while cat is 40%. What this imply is that Cats only has greater chances of being adopted than Dogs only in the first seven days.

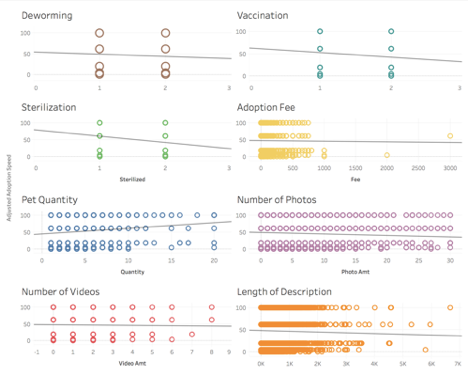
To verify the above we apply the remaining 30% of the test data to the model. The result is shown below:



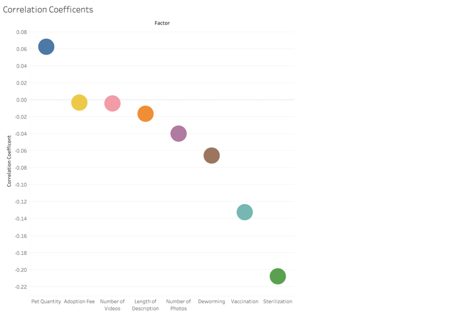
The above shows that when adoption speed is less than or equal to 1 (week 1: 0-7 days) there is 55% chance of adopting Cats and 45% chance of Dogs. However, when the adoption speed is greater than 1 (more than one week), you have 55% chance of adopting a Dog and 45% chance of adopting Cats. These further stresses the result of the train data. We can conclude that if an animal is adopted quickly it is more likely to be a cat, but of the animals which are adopted after a week the proportion shifted in favor of dogs.

It is important when choosing factors while attempting to improve a process that you focus on factors that are actually able to be controlled. For this dataset the factors which can be controlled once the pet is available for adoption are vaccinations, deworming, sterilization, number of pets in the profile, adoption fee, quantity of photos and videos, and the length of the description. When examining this dataset, the nature of the Adoption Speed metric complicates analysis, because of this trying to convert the metric back into total days is a useful method for getting more meaningful analysis. There was no way to get a perfect conversion so in order to estimate the conversion was done by converting the Adoption Speed metric into the average for the range of each bin and using 100 as the conversion for an Adoption Speed of 4 so as to prevent overestimating.

In order to analyze correlation of the controllable factors against the adjusted adoption speed all of the factors were plotted against the adjusted adoption speed and then the trend for each factor was plotted.



Upon immediate evaluation it is clear that though the adjustment to the adoption speed metric improves the evaluation it is by no means a perfect solution and this fact must be kept in mind when determining the value of the results. All but one of the factors showed a negative correlation, this means that having an animal dewormed, vaccinated, or sterilized, having more photos and videos, and a longer description are all correlated with faster adoption times. Quantity of pets in the listing, however is correlated with a longer adoption time. A larger adoption fee is correlated with a shorter adoption time, but it appears to be heavily influenced by a few outliers. Exploring deeper the correlation coefficient for the individual factors shows that no factor is especially strongly correlated in either direction.



Looking at the correlation coefficients the strongest is only -.21 which is not especially strong. It is possible that these numbers are deflated by the fact that we do not have exact data for the days of adoption which limit how well the data fits a model.

Attempting to analyze the data for this project was significantly hampered by the limited nature of the adoption speed factor. By not allowing for more precise evaluations the value of the resulting analysis was subsequently diminished, however this does not mean that there were no meaningful results to be drawn. The correlation results were not overwhelmingly conclusive, but the results (especially the vaccination and sterilization ones) were strong enough that it would be worth investigating further, particularly if the exact adoption speed data could be acquired. The classification model showed a noticeable difference between the adoption rates by animal between the first week and later adoption which warrants further exploration in attempt to determine the cause behind the phenomenon. While the analysis was not wholly conclusive it did provide avenues for further exploration down the road.

**References**

Resources – Koret Shelter Medicine Program. (n.d.). Retrieved from

https://www.sheltermedicine.com/library/resources/?utf8=u2713&search[slug]=adoption-driven-capacity