Deanna Aburaad
Amanuel Alemu
June 8th, 2016

Final Report

Data:

There are a little over 43,800 instances in the dataset we obtained from Austin animal shelters. (https://data.austintexas.gov/). We then pre-processed that data as follows.

- No outcome subtypes were included. These are things that describe the nature of the outcome itself. E.g. reason for death. These were excluded because it is dependent on the target we want, and it is not an attribute we care about for this task. That is, if we can predict that an animal has not been euthanized, we do not want to predict why, and knowing why it was euthanized would throw off our data.
- All mixed breeds (i.e. "bat mix") were simplified to their main breed
- Spayed/neutered male/female attribute was split into two attributes of:
    - Intact (spayed/neutered), indicated by 0 for intact and 1 for altered; and
    - Sex, indicated by male or female
- Age converted to weeks (continuous)
- Named and unnamed animals were indicated as such by a 1 or 0, respectively.
- Dates were discarded, as they did not indicate the length of the shelter stay, only when the outcome was recorded.
- Filtered out instances with outcomes of transfer, missing, relocated, and return to owner, resulting in 21,746 examples to train with outcomes of adopt, euthanasia, die, and disposal. This is because those are not outcomes that are at all related to an animal's characteristics, which is what we are looking into.

Methods and Testing:

We used Orange (http://orange.biolab.si/) to test our altered file, choosing to use a classification tree. We compared the results from our previous progress report with those of our current model as a way to measure our improvement since then, after filtering out some of the more extraneous data that was included in the original austin dataset.

When testing, we found that the updated data in the test file we had originally saved was not comparable to the data we had used to train the decision tree, since Austin changed the features they kept track of between those particular datasets. Specifically, the dataset we used for training comprised of intakes and outcomes from October of 2013 up to now, and the test data we had saved was the intake and outcome data from prior to 2012. However, it seems that around 2013, Austin changed what included in those datasets. Among those changes were the inclusion of animals other than cats and dogs, the names of the animals (when possible), and some changes in how they recorded the breeds of the animals. While the first two changes were manageable as far as still using the dataset for testing, the third change in particular made it difficult to use since a decent number of the breed attributes did not match up across datasets. Due to these discrepancies, we decided to split our training data 70/30, so that we'd now train on 70% and test on the remaining 30%. SInce we had a large number of instances on record,

Deanna Aburaad
Amanuel Alemu
June 8th, 2016

even after we filtered out the instances we didn't care about, we figured that splitting it would be our best bet.

Analysis and Future Work:

[1]The Humane Society points out the fluctuations in cost per capita and animal intake/euthanization rates across the United States, in comparison with the average:

> *On average, communities in the United States:*
> - *Spend approximately $8 per capita for animal shelters*
> - *Handle on average around 30 animals per 1,000 people*
> - *Euthanize about 12.5 animals per 1,000 people*

For example, in Multnomah County, Oregon, the community spends more per capita at $16 but euthanizes only 7 animals per 1,000 people. This extra money might go towards extra enforcement, or towards spaying and neutering the animals that come in, resulting in fewer animals being euthanized.

Our results indicate that spayed/neutered animals 1.5 years old and younger are more likely to be adopted, which supports the argument that shelters should put their funds towards spaying and neutering animals that otherwise would have a high chance of being adopted.

However, since our data is taken from only one no-kill shelter in Austin, Texas, it is certain that there is bias in our results. Animals were only euthanized if they were suffering or aggressive, whereas other shelters euthanize their animals due to overcrowding. The later of these shelters might find enforcement to be more cost effective in saving animals than spaying or neutering.

In conclusion, our findings are most relevant when compared with other no-kill shelters. Considering the lack of data nationwide, let alone standardized data, on animal outcomes, future work should be done to gather more data on important/missing attributes (length of stay at shelter, weight of animal, cost to maintain animals, health status of the animal, etc.). Work should also be done to monitor the work being done at the roughly [2]13,600 independent shelters around the US and to accurately gather national statistics, instead of having to rely on estimates.

---

[1]

http://www.humanesociety.org/animal_community/resources/timelines/animal_sheltering_trends.html?referrer=https://www.google.com/?referrer=http://www.humanesociety.org/animal_community/resources/timelines/animal_sheltering_trends.html

[2]

http://www.aspca.org/animal-homelessness/shelter-intake-and-surrender/pet-statistics