Furry Futures: Predictive Analytics for Adoption Outcomes in No-Kill Shelters

Tess Anderson

tanderson4@bellarmine.edu

January 16, 2024

1. Executive Summary

The focus of this project is predicting trends in animal adoptions, specifically cats, using predictive analytics. I aim to develop a predictive model to predict the outcome of cats' time in the shelter based on features such as age, gender, color, and breed. Some of the most common outcomes include adoption, transfer, euthanasia, death, and foster. This approach to analyzing these data will result in valuable insights that can inform the shelter of whether a cat is likely to be adopted, or if the focus should be on foster placement or transferring the animal. This project utilizes a dataset sourced from kaggle.com which contains data on over 28,000 cats. In this analysis, I will explore and utilize predictive models beyond linear and logistic regression such as a Random Forest Classifier, Support Vector Machines (SVM), and Decision Trees – parameter fine tuning will also be employed to generate the best possible results for these data.

2. Project Idea

In this project, I aim to predict trends in cat adoptions using a large dataset from the Austin Animal Center, a large no-kill animal center in Austin, Texas. The dataset contains information on animals that entered the shelter including several relevant predictors, as well as outcome which is the classification I aim to predict. Using a Random Forest model, Decision Tree model, or Support Vector Machine, I will analyze the data to identify patterns and factors which function as valuable predictors of cats' outcomes. I will use Python to preprocess the data, perform exploratory data analysis, and train and test the models. I will use Tableau to generate visualizations for EDA and to communicate and present my findings.

3. Background

Stray animals, in particular, cats, are a challenge in society that animal shelters set out to address. Stray animals suffer – exposed to the elements, often without the instincts to live on their own and protect themselves – and present a public health and safety for humans as well. [1] It has been observed in recent years that the numbers of dogs entering shelters are decreasing, while the numbers of unwanted cats are increasing. [2] Animal shelters attempt to address the stray animal populations, however many shelters are overwhelmed, understaffed, and insufficiently funded. This analysis focuses on the Austin Animal Center in Austin, Texas which aims to provide necessary food, water, shelter, and standard veterinary care for animals in need. Since 2010, the shelter has been following the guidelines of the City of Austin's No-Kill implementation plan and has saved more than 90% of the approximately 20,000 animals entering the shelter each year. [3] As a no-kill shelter, it is important that the shelter

uses its finite resources wisely to effectively care for as many animals as possible. Allocating resources efficiently includes several factors such as staffing, medical care, and space allocation.

The Austin Animal Center is the largest no-kill animal shelter in the United States and is already an outstanding animal shelter that has effective policies and guidelines that guide decision-making. They also receive great support and funding from their community through donations and through allocation of government funds. However, data science and analysis can further improve and inform their policies and practices to ensure that their resources are being used most efficiently. Using predictive analytics, animals' individual needs can be determined and planned for effectively to make the best use of the limited resources available to the shelter. This analysis can improve adoption rates, reduce the length of stays at the shelter, and enhance animal welfare and care by informing strategic decision-making.

More specifically, this research will result in insights that can help identify the likely challenges and needs based on the individual features of each cat that comes through the shelter. By identifying the cats who are most likely to get adopted easily, the shelter can include the most adoptable pets at adoption events to decrease the length of stays and increase the success of events such as these. By identifying the cats who are most likely to need foster care, the shelter can save valuable time and space by beginning that search early. By identifying the cats who are most likely to be euthanized, the shelter can make humane and cost effective decisions that benefit the shelter, the suffering cat, and all other animals in the shelter.

The dataset I will use for this analysis is sourced from Kaggle.com and contains 37 columns and over 28,000 rows. The observations are cats that entered the Austin Animal Services system between 10/1/2013 and 02/01/2018. The dataset contains shelter outcomes of cats, as well as many different features such as breed, age, sex, whether the cat is spayed/neutered, color, and coat pattern and type. These features will inform the predictions generated by the predictive models – the models will be trained on a portion of the data to analyze the relationships that exist in the dataset. Theoretically, these features are significant predictors of outcome as adopters make decisions based on their preferences (sex, breed, color, coat) and by considering practical elements such as their age and whether the cat is spayed or neutered. There are many different outcomes included in the dataset which we aim to predict – some of the most common outcome types are adoption, transfer, euthanasia, died, and spay-neuter-release. The model will be designed to predict the outcome based on these features which theoretically will have a significant impact on outcome and will be valuable predictors.

4. Modeling

Three potential models for predicting the outcomes are Random Forest Classifier, Support Vector Machines (SVM), and Decision Trees. Random Forest Classifier is well-suited for classification tasks such as these and handle non-linearity well. Additionally, this model is appropriate for datasets which contain a mix of categorical and numerical features such as this one. Support Vector Machines (SVM) are useful in handling complex decision boundaries and are well-suited for multi-class classification tasks such as this one. Decision Trees are similar to Random Forest models – in fact, Random Forest is made of many Decision Trees. I am interested in exploring how these two models compare in performance with their similar concepts and approaches to classification. Logistic Regression will be used as a baseline to aid in performance comparison across all models.

5. Tools

This project will mostly be executed using Python in Jupyter notebooks. Python and its wide range of libraries allows me to import, manipulate, clean and preprocess, and explore the data, as well as employ pre-trained machine learning models to generate a predictive model through training and testing. Some of the libraries I will leverage to carry out this analysis are Pandas, NumPy, SciKit-Learn, and potentially Seaborn and Matplotlib to generate visualizations. Though some simple visualizations will likely be generated using Python during the EDA phase of this analysis, most of my visualizations will be created using Tableau. I will use this visualization tool to create more sophisticated representations of the relationships and distributions that exist in the dataset, as well as present my findings in a dashboard. I chose to use Python as the programming language because I find it to be more flexible than other languages like R, and the abundance of helpful packages that are available for free will help with every step of this analysis from preprocessing to EDA to modeling. I chose to use Tableau as a visualization tool because I have more experience working with Tableau than other tools such as Power BI – both, however, are powerful and user-friendly tools that help generate useful representations of the insights I will uncover through this analysis.

6. Conclusion

In conclusion, this project sets out to transform and improve the approach of animal shelters to finding loving homes for stray and vulnerable pets by using predictive analytics to forecast outcomes for animals based on their individual features. Through the implementation of predictive models like Random Forest Classifier, Decision Tree Classifier, and SVM, shelters can move beyond tradition decision-making and embrace a data-driven approach when creating policies, strategies, and interventions for animals in need. The tools and methodologies employed in this research

have been carefully selected to best accomplish the goals of this project and to appropriately handle the diverse features of the data. The insights gleaned from this project can inform efficient resource allocation and improve strategic decision-making to ultimately maximize the number of animals who achieve the ideal outcome for a shelter pet – finding a forever home.

7. References

[1] Aaronschlegel. (2018, February). Austin Animal Center Shelter Outcomes. Retrieved January 12, 2024 from https://www.kaggle.com/datasets/aaronschlegel/austin-animal-center-shelter-outcomes-and/data.

[2] Turner, P., Berry, J., & Macdonald, S. (2012). Animal shelters and animal welfare: raising the bar. *The Canadian veterinary journal = La revue veterinaire canadienne*, *53*(8), 893–896.

[3] Morris, K. N., Wolf, J. L., & Gies, D. L. (2011). Trends in intake and outcome data for animal shelters in Colorado, 2000 to 2007. *Journal of the American Veterinary Medical Association*, *238*(3), 329-336.

[4] “No-Kill Plan.” AustinTexas.Gov, www.austintexas.gov/page/no-kill-plan. Accessed 16 Jan. 2024.