Capstone Project Proposal: Finding Pets a Home

**Problem Statement**

According to The American Society for the Prevention of Cruelty to Animals, US animal shelters takes in approximately 6.5 million companion animals nationwide every year and 1.5 million shelter animals would be euthanized every year; that is 1 in 5 animals entering shelters never leave! To combat this, I propose to investigate the use machine learning to decrease the rate of euthanization by creating a smart transferring system between pet adoption agencies to improve outcomes for shelter animals. To achieve this the following questions would be explored. What constitutes the decision to such high euthanize animals? Are there certain key features that predispose the shelter animals to euthanization? What are the keys characteristics of entering animals that lead to easy adoption? Are these characteristics different by region of the United States? This project would attempt at shedding light to these questions and gaining understanding of the facing moral problem.

**Who would care?**

The public has always had compassion for companion animals. Animals have integrated into many as a part of their daily lives. They act as companions during emotionally difficult times and provide a sense of joy in day to day life. Furthermore dogs have been assimilated into many industries trained as support dogs to provide services from medical alerts for diabetic patients to drug detection on suspicious cargo crossing the border. With the long history between human and animals, the public has felt the need provide moral and sensible support for helpless animals. The problem would also provide awareness to animal loving individuals and inspire some to support animal shelters and other animal rights organization.

Animals shelters can use the discovery to identify high risk sheltered animals and provide extra aids to improve outcomes. This could remedy long sheltered times, freeing up sheltered spaces for in-needed animals and reallocating resources to other areas of the shelter. Financially speaking this would provide monetary remediation for animal shelters by reducing operating expenses in this area and provide funding to improve conditions in the shelter or other parts of the organization. Ultimately this would help to provide a closer step to their mission of providing care and rehabilitation to animals.

Animal rights organizations has long fought for the rights of animals living conditions, abuse and euthanization. Improvement in animals outcome aligns with their mission and their philosophy

**Data:**

According to the American Humane, animal shelters and other care agencies are not required to keep data on animals, hence the data available on the web are difficult to acquire and must be scoured to find. Currently 5 datasets from different regions of the US were found dating as far back as 2013 with feature vectors such as ‘type of animal’, ‘breed’, ‘color’, ‘outcome’, ‘entering date’, etc.

<https://data.austintexas.gov/Health-and-Community-Services/Austin-Animal-Center-Outcomes/9t4d-g238>

<https://data.bloomington.in.gov/dataset/adopted-animals>

<https://www.dallasopendata.com/City-Services/Dallas-Animal-Shelter-Data/7h2m-3um5>

<https://data.louisvilleky.gov/dataset/animal-service-intake-and-outcome>

<https://data.sonomacounty.ca.gov/Government/Animal-Shelter-Intake-and-Outcome/924a-vesw>

**Approaches to solving this problem:**

Python numpy and pandas packages would be used for data wrangling. All the datasets would be merged and cleaned. Inconsistencies and missing data would be checked and eliminated through logical procedures.

Exploratory analysis would be done through matplotlib and seaborn packages to provide a visual understanding of the underlying data, illuminating interesting discoveries along the way. Correlations and hypothesis testing would be done for sheltered animals in different regions and among different animals themselves. Numpy and scipy.stats would be the main packages to explore the statistics.

Mostly classification models would be used on the dataset to predict the outcomes for animals and as a result sci-kit learn would be the main package for the machine learning part of the project

Deliverables:

The code for the project and report would be available on github. A short powerpoint would also be available for project as well.