COMPSCI 526 Project Proposal

Title: Data-Driven Discoveries in Pet Adoption Patterns

Project Category: social science

Project Members:
Yushan Shi ys468
Yuqi Su ys457
Yurui Wu yw695
Mingxin Liu ml754

Problem Statement

Millions of stray animals suffer on the streets or are euthanized in shelters daily. Our project aims to unfold factors influencing animal adoption speeds by analyzing a comprehensive dataset of adopted pet profiles. Identifying hidden patterns that impact adoption speed can help shelters create more appealing online profiles, improving adoption rates and benefit global animal welfare.

Challenges

Our analysis faces several complex challenges. The dataset contains high-cardinality categorical variables, such as breed and color combinations, that need careful handling to extract meaningful insights. Identifying interesting patterns that are not widely recognized while avoiding misleading correlations requires thorough research and analysis. We must also address potential biases in the data and interpret feature interactions to ensure accurate, meaningful results.

Dataset

We will use the PetFinder.my Adoption Prediction dataset, available on Kaggle. It contains detailed information on pet profiles, including: Pet characteristics, health conditions, and the adoption speed. Our strategy involves thorough EDA, detailed preprocessing to handle common issues, such as missing values and encoding categorical variables. The dataset also contains longer text descriptions and images that we can potentially explore and gain other meaningful information on.

Method/Algorithm

Our analysis will use following steps to extract insights:

1. *Exploratory Data Analysis:*
   i. Visualize distributions and relationships using histograms, box plots, and correlation heatmaps.
   ii. Conduct statistical tests (e.g ANOVA) to identify significant relationships between features and adoption speed.
2. *Feature Importance Analysis:*

    i. Train ensemble models, such as Random Forest, to obtain feature importances.

    ii. Use LIME/SHAP/SHARP values for more detailed explanations for individual instances.

3. *Advanced Visualizations:*

    i. Create Partial Dependence Plots to visualize how individual features affect our target variables.

    ii. Visualize simple decision trees to understand key splits in the adoption process.

4. *Insight Extraction:*

    i. Identify what could be the shocking patterns for adopting pets.

    ii. Further explain the causality between certain pets' features and their adoption speed.

We will implement the methods above with Python, and the goal is to predict adoption speed, but more importantly, we want to understand the possible factors that influence the adoption process.

Literature Review

Research on pet overpopulation and adoption highlights the need for strategies encouraging adoption from shelters, as millions of animals are abandoned or euthanized annually in the U.S. (Frank, 2008)[i]. Studies show that factors such as breed, size, age, and personality significantly influence adoption rates. Smaller dogs and exotic breeds tend to be adopted faster than others, while traits like friendliness, playfulness, and affection are crucial in adoption decisions (Diesel et al., 2007; Gourkow, 2001)[ii] [iii]. Age also plays a key role, with younger dogs and adult cats adopted more quickly than older animals (Normando et al., 2006; Janke et al., 2017)[iv] [v]. Male cats and long-haired breeds tend to have shorter stays in shelters, though behavioral and health issues can hinder adoption (Kay et al., 2018)[vi]. Other factors, such as population density around shelters and pet appearance, including coat color, also influence adoption rates (Miller et al., 2019)[vii]. While appearance is a primary factor for dog adoption, behavior is more critical for cats (Weiss et al., 2012)[viii]. Common research methods include textual analysis, regression models, decision trees, and random forests to analyze these trends (Zadeh et al., 2022)[ix].

Evaluation

Quantitative Measures includes F1 score, Accuracy, AUC-ROC Curve etc. Additionally, ANOVA/Chi-Square Tests and Correlation Analysis.

As for Qualitative evaluation, by using Random Forest and SHAP/LIME values, we will generate explanations for individual predictions and assess the overall importance of specific features such as breed, age, and health condition. Partial dependence plots allow us to visualize how specific features (e.g., pet age or health condition) impact adoption speed, providing interpretable insights of factors influencing adoption speed.

[i] Frank, J.M. (2008). Attitudes and Perceptions Regarding Pet Adoption.

[ii] Diesel, G., Pfeiffer, D. U., & Brodbelt, D. (2008). Factors affecting the success of rehoming dogs in the UK during 2005. *Preventive veterinary medicine*, *84*(3-4), 228-241.

[iii] Gourkow, N. (2001). *Factors affecting the welfare and adoption rate of cats in an animal shelter* (Doctoral dissertation, University of British Columbia).

[iv] Normando, S., Stefanini, C., Meers, L., Adamelli, S., Coultis, D., & Bono, G. (2006). Some factors influencing adoption of sheltered dogs. *Anthrozoös*, *19*(3), 211-224.

[v] Janke, N., Berke, O., Flockhart, T., Bateman, S., & Coe, J. B. (2017). Risk factors affecting length of stay of cats in an animal shelter: A case study at the Guelph Humane Society, 2011–2016. *Preventive Veterinary Medicine*, *148*, 44-48.

[vi] Kay, A., Coe, J. B., Young, I., & Pearl, D. (2018). Factors influencing time to adoption for dogs in a provincial shelter system in Canada. *Journal of Applied Animal Welfare Science*, *21*(4), 375-388.

[vii] Miller, H., Ward, M., & Beatty, J. A. (2019). Population characteristics of cats adopted from an urban cat shelter and the influence of physical traits and reason for surrender on length of stay. *Animals*, *9*(11), 940.

[viii] Weiss, E., Miller, K., Mohan-Gibbons, H., & Vela, C. (2012). Why did you choose this pet?: Adopters and pet selection preferences in five animal shelters in the United States. *Animals*, *2*(2), 144-159.

[ix] Zadeh, A., Combs, K., Burkey, B., Dop, J., Duffy, K., & Nosoudi, N. (2022). Pet analytics: Predicting adoption speed of pets from their online profiles. *Expert systems with applications*, *204*, 117596.