**Petfinder Pawpularity**

Kayla Thompson

DSC 680 - Fall 2021

<https://github.com/kaymaethomp1>

**Abstract**

Adopting a pet has been a choice in my life that has had a major impact on my happiness. The aim of this project is to create a solution for more pets to find more families. In this project I will use a variety of machine learning techniques in order to create a model that will predict the popularity of a pet’s adoption page based on various characteristics of their photo. An intended outcome of this project will be the ability for an adoption agency to optimize the pet’s popularity score in order to lead to a quicker adoption.

**Problem Statement**

Speaking to my own personal experiences, adopting a pet has completely changed my life. I am fortunate that the process I completed in order to find my pet was relatively simple and straightforward, but this isn’t always the case. Finding the right person the right pet is a game of chance and is largely dependent on the amount of exposure an adoptable pet receives online. According to research, owning a pet includes benefits like a higher survival rate for heart attacks, lower stress levels, and emotional support [1]. In return the pet receives a loving home with endless attention and great care. However, adoption isn’t always as simple as it seems. With limited resources and stretched capacity, rescue, and adoption organizations are doing all that they can to help as many animals as possible. In this project, we will aim to create a tool for those agencies in order to optimize a pet’s adoption profile. Pictures are one of the most important pieces to a pet’s profile and the algorithm built in this project will help adoption agencies optimize the photos selected for the pet’s profile in order to increase a pet’s popularity (“pawpularity”).

**Methods and Results**

To begin work on this project, I started with exploratory data analysis. The goal of this analysis is to get a fundamental understanding of what the data is comprised of and how the data is structured. The results of this analysis will also inform what data clean-up is necessary before proceeding. After looking at the data the “ID” column is the only column that is not necessary and therefore I drop this column from the pandas data frame. After this column is dropped, the next step in my analysis was to calculate the mean, standard deviation, mode, variance, and tail of the target variable (‘pawpularity’) using functions from thinkstats2. The mean is 38.039, the Standard Deviation is: 20.59, the Spread is: 424.03, the Mode is 28 and 30, and the Tail is 15, 70, 20, 20, 30. I also created a histogram for the target variable (figure 1). What was most important about these results is that the data has a fairly large spread but is a normal distribution. The data is slightly skewed toward the lower values. The next step in the exploratory data analysis is to use matplotlib and seaborn to calculate and plot the correlations between each variable. The goal of this analysis to use build a model that predicts the pawpularity score based on the provided attributes, strong correlations would be an indication of what variables are most important for the model. The output of the correlation is shown in the appendix (figure 2). The results of this correlation did not show any strong relationships between any variable and pawpularity. This to me indicates that each variable is equally important to include in the model. Once I come to this conclusion, the exploratory data analysis is complete. The next step, is to begin building and testing my models. Before building any models, I first use sklearn to divide my data into a test set and a training set. Once this is done I begin work on the models. For this project I have decided to test a Random Forest model and a linear regression model. I used sklearn to build the random forest. Once it was built, I ran the test data set through the model in order to then calculate the root mean square error. Root mean square error in this case, measures the differences between the actual pawpularity scores and the pawpularity scores predicted by the model. The root mean square error for the random forest model is 20.897. Next, I use sklearn to build the linear regression model using the training data set. Similar to the random forest model, I also run the test set through the model and I calculate the root mean square error. The root mean square error for the linear regression is 20.624.

**Conclusion and Additional Questions**

Based on the results of the model building and testing, the more appropriate model for this problem is the linear regression model. The linear regression has the lowest root mean square error which means that the difference between the actual values and the predicted values is smaller in the case of the linear regression model. However, based on the exploratory data analysis and the high root square mean error in both models, I feel that we do not have all of the relevant variable and information for this question. There is likely a variety of external factors at play with a pet’s pawpularity score that aren’t accounted for in the data. A pet’s age, breed, coloring, location and temperament are all factors that could make a pet more or less popular on Petfinder. Knowing the data is missing these pieces, knowing that none of the variables correlated with pawpularity score and knowing the models do not perform well leaves the conclusions of this project largely inconclusive. Future work should consider these other variables in order to take a more holistic view at an adoptable pet’s profile and ultimately pawpularity score.

**Appendix**

Figure 1

Chart, histogram

Description automatically generated

Figure 2

Chart

Description automatically generated

**Sources**:

[1] Ari Notis December 20, 2017. (2019, July 8). *15 amazing benefits of adopting a pet*. Best Life. Retrieved October 31, 2021, from https://bestlifeonline.com/adopt-a-pet/.

Eller, K. (2018, June 27). *The health benefits of adopting a dog*. health enews. Retrieved October 31, 2021, from https://www.ahchealthenews.com/2017/04/11/health-benefits-adopting-dog/.

Harris, H. (2007, October 23). *Using petfinder 101*. HowStuffWorks. Retrieved October 31, 2021, from https://animals.howstuffworks.com/pets/petfinder.htm.

*How adoption works*. Animal Humane Society. (n.d.). Retrieved October 31, 2021, from https://www.animalhumanesociety.org/adoption/how-adoption-works.

Lingel, G. (2020, December 12). *Animal rescue: What it is and why you should get involved*. Sentient Media. Retrieved October 31, 2021, from https://sentientmedia.org/animal-rescue/.

*Petfinder.my - pawpularity contest*. Kaggle. (n.d.). Retrieved October 31, 2021, from https://www.kaggle.com/c/petfinder-pawpularity-score/data.

*Preparing to adopt*. Animal Humane Society. (n.d.). Retrieved October 31, 2021, from https://www.animalhumanesociety.org/adoption/preparing-adopt.

Robinson, L. (2021, July 19). *The health and mood-boosting benefits of pets*. HelpGuide.org. Retrieved October 31, 2021, from https://www.helpguide.org/articles/mental-health/mood-boosting-power-of-dogs.htm.

Thomas, T., & Kime, P. submitted by J. (2020, January 7). *Give love, get love-the emotional benefits of adopting a pet*. Richland Source. Retrieved October 31, 2021, from https://www.richlandsource.com/thrive/give-love-get-love-the-emotional-benefits-of-adopting-a-pet/article\_b6beede2-3179-11ea-aee4-07e002b06bda.html.

*Top reasons to adopt a pet*. The Humane Society of the United States. (n.d.). Retrieved October 31, 2021, from https://www.humanesociety.org/resources/top-reasons-adopt-pet.