

Capstone Project Proposal

Kaggle Challenge: PetFinder.my Adoption Prediction

Mahesh Babu Gorantla Feb 18, 2019

Domain Background

Millions of stray animals suffer on the streets or are euthanized in animal shelters every day around the world. If at all we can find those stray animals a forever home, many precious lives can be given a new life that they truly deserve and in turn we will see more happier families.

[PetFinder.my](#) has been Malaysia's leading animal welfare platform since 2008, with a database of more than 150,000 animals. PetFinder collaborates closely with animal lovers, media, corporations, and global organizations to improve animal welfare.

Problem Statement

The Pet Adoption Prediction is a multi-class classification problem. In this problem, I need to develop algorithms to predict the adoptability of pets - specifically, how quickly is a pet adopted based on its features (which have been explained in the section below).

It is a classification problem because a pet's adoption speed is classified into 5 categories.

A relevant potential solution is to use a few classification algorithms such as

- k-Means Clustering (Baseline model)
- DBSCAN
- Random Forests
- XGBoost

And, choose the one with best **cross_validation** score on the training dataset.

Datasets and Inputs

In this problem, I will predict the speed at which a pet is adopted, based on the pet's listing on PetFinder. Sometimes a profile represents a group of pets. In this case, the speed of adoption is determined by the speed at which all of the pets are adopted. The data included text, tabular, and image data.

File Descriptions

- **train.csv** - Tabular/text data for the training set
- **test.csv** - Tabular/text data for the test set
- **sample_submission.csv** - A sample submission file in the correct format
- **breed_labels.csv** - Contains Type, and BreedName for each BreedID. Type 1 is **dog**, 2 is **cat**.
- **color_labels.csv** - Contains ColorName for each **ColorID**
- **state_labels.csv** - Contains StateName for each **StateID**

Data Fields

- **PetID** - Unique hash ID of pet profile
- **AdoptionSpeed** - Categorical speed of adoption. Lower is faster. **This is the value to predict.** See below section for more info.
- **Type** - Type of animal (1 = Dog, 2 = Cat)
- **Name** - Name of pet (Empty if not named)
- **Age** - Age of pet when listed, in months
- **Breed1** - Primary breed of pet (Refer to BreedLabels dictionary)
- **Breed2** - Secondary breed of pet, if pet is of mixed breed (Refer to BreedLabels dictionary)
- **Gender** - Gender of pet (1 = **Male**, 2 = **Female**, 3 = **Mixed**, if profile represents group of pets)
- **Color1** - Color 1 of pet (Refer to ColorLabels dictionary)
- **Color2** - Color 2 of pet (Refer to ColorLabels dictionary)
- **Color3** - Color 3 of pet (Refer to ColorLabels dictionary)
- **MaturitySize** - Size at maturity (1 = **Small**, 2 = **Medium**, 3 = **Large**, 4 = **Extra Large**, 0 = **Not Specified**)
- **FurLength** - Fur length (1 = **Short**, 2 = **Medium**, 3 = **Long**, 0 = **Not Specified**)
- **Vaccinated** - Pet has been vaccinated (1 = **Yes**, 2 = **No**, 3 = **Not Sure**)
- **Dewormed** - Pet has been dewormed (1 = **Yes**, 2 = **No**, 3 = **Not Sure**)
- **Sterilized** - Pet has been **spayed** / **neutered** (1 = **Yes**, 2 = **No**, 3 = **Not Sure**)
- **Health** - Health Condition (1 = **Healthy**, 2 = **Minor Injury**, 3 = **Serious Injury**, 0 = **Not Specified**)
- **Quantity** - Number of pets represented in profile Fee - Adoption fee (0 = **Free**)
- **State** - State location in Malaysia (Refer to StateLabels dictionary)
- **RescuerID** - Unique hash ID of rescuer
- **VideoAmt** - Total uploaded videos for this pet
- **PhotoAmt** - Total uploaded photos for this pet
- **Description** - Profile write-up for this pet. The primary language used is English, with some in Malay or Chinese.

AdoptionSpeed

Contestants are required to predict this value. The value is determined by how quickly, if at all, a pet is adopted. The values are determined in the following way:

- 0 - Pet was adopted on the same day as it was listed.
- 1 - Pet was adopted between 1 and 7 days (1st week) after being listed.
- 2 - Pet was adopted between 8 and 30 days (1st month) after being listed.
- 3 - Pet was adopted between 31 and 90 days (2nd & 3rd month) after being listed.
- 4 - No adoption after 100 days of being listed. (There are no pets in this dataset that waited between 90 and 100 days)

Solution Statement

Benchmark Model

Evaluation Metrics

Project Design