Grenze International Journal of Engineering and Technology, Jan Issue

Pet Adoption Speed Prediction

B. Akhil[1], B. Shashank[2], Eliganti Ramalakshmi[3] and Dr.T Prathima[4]
[1-4]Chaitanya Bharathi Institute of Technology Hyderabad, India.
Email: Akhilbhashyam23@gmail.com, bhimabithinshashank18@gmail.com, eramalakshmi_it@cbit.ac.in, tprathima_it@cbit.accc.in

***Abstract*— Many stray animals throughout the world unfortunately do not get the opportunity to find the loving home that they deserve. In this classification task, we look to develop an algorithm to predict the speed of pet adoptions. The input to our algorithm includes animal type (dog or cat), breed, gender, colour, profile picture online and the descriptions about the pet, etc. We then use both traditional machine learning techniques (Logistic Regression, Naive Bayes, SVM, Decision Tree, Random Forest and Gradient Boosting) and neural networks (fully connected neural networks and long short-term memory Model) to predict the adoption rate. In particular, we build feature vectors using information extracted from description scripts and feed them into neural network models. We are hoping to examine the results to develop strategies to help improve the overall adoption rate (i.e. what features lead to faster adoption).**

Every year, 3.3 million canines visit animal shelters, out of a total population of 200 million. Only 2% to 17% of pets are returned to their owners. The remaining animals are euthanized due to a shortage of room in shelters (killed). The present procedures for adapting or finding a pet are inefficient and haphazard. People disseminated leaflets to the general public and spread the word to others in the vicinity of the pet's disappearance. When individuals print fliers, they waste paper and money since there is no good impact. People also share their tales on social media sites like Instagram and Facebook. There are also several fraudsters who attempt to fraudulently claim the incentive for returning the pet to its legitimate owner. We want to provide an analysis on how fast pets(cats and dogs) can be adopted based on various factors like their health conditions, age, colour, breed etc.

I. INTRODUCTION

Animal adoption is the procedure through which a human takes in an animal, most often a cat or a dog. Animal control takes these lost or abandoned animals and places them in animal shelters. Pets that have been in shelters for an extended period of time are killed in order to alleviate the overpopulation of undesirable animals. Animal adoption has seen a solid trend in the society for decades. We have come very far with the way we manage abandoned animals and strays to rehabilitate them. While organizations like the Blue Cross of India (BCI) have operated since the year 1959, with the advancement of technology, we now have the system of online pet adoption where databases of pets taken in by numerous rescue groups and animal shelters are maintained and are openly available for access and utilization by the public. Further, a lot of innovative work has being going into the field of animal welfare.

Some people euthanize pets because of fatal and irrecoverable injuries or illnesses. Others do so even for treatable health problems that they either are unable to financially provide for or are not willing to. Without doubt, many veterinarians consider euthanasia to be a completely unethical procedure. Some other people let go of their pets, into the wild perhaps, or abandon them with the hope that they will be able to care for themselves or that someone

Grenze ID: 01.GIJET.10.1.303

will find or adopt them. In reality, most of these pets that are released by their owners perish to harsh weather, hunger, traffic or easily treatable health issues. In addition to this, homes cannot be always found in adoption centres and unless the centre has a no-kill policy, excess animals are also euthanized to make room for new pets. It was estimated by The Humane Society of the US, that every year, 2.4 million adoptable, healthy dogs and cats are euthanized in the United States of America because of unavailability of homes.

A. Objectives

- To understand what features play a major role in determining the speed of adoption of a pet.
- To determine if a pet would be adopted or not based on the description and images of dogs and cats provided.
- To use H2O algorithm to determine the speed of adoption.

II. Literature Survey

Paul C. Bartlett, stated that "In 2003, of the 140,653 dogs that were discharged from Michigan shelters in 2003, it was found that 28% were taken in for adoption and 40% had to undergo euthanization; this implied that the annual euthanasia rate was 2.6% for the estimated 2003 Michigan population of dogs. The annual rate of euthanasia for cats in 2003 was much higher than that of dogs (3.1% and 2.6%). Of the 134,405 cats discharged from Michigan shelters in 2003, only 24% were adopted and 57% were euthanized. For both dogs and cats, altered animals had much lower rates of euthanasia than unaltered animals. Among government and private owned animal shelters, private shelters had 20% of their dogs euthanized and 29% of their cats compared to 30% of dogs and 50% of cats euthanized in facilities owned by the government. Additionally, the results indicated that the larger the size of the shelter, the higher the rate of euthanasia for both dogs and cats regardless of ownership. [1]

Brown W.P analysed the effects of age, sex , breed designation, coat colour, and coat patterns on the length of stay (LOS) for cats in a no-kill shelter; and also compared the physical characteristics and LOS of cats adopted at a no-kill shelter and in traditional shelters. His article indicates feline physical characteristics influence LOS, likelihood of adoption, or risk of euthanasia at shelters of all kinds and recommend exploring strategies to increase the adoptability of less desirable cats and kittens. [2]

Rachael Mozes evaluated the factors associated with euthanasia in an animal shelter in Kitchener-Waterloo, Ontario, Canada. He concluded that 53% of admitted cats were euthanized. The adoption policies used in animal shelters have a great effect on the total number of adoptions and donations that occur. [3]

Lord .E established a stochastic model to simulate the expected costs, revenues, and net income of a hypothetical animal shelter for various alternative management strategies, based on US conditions. [4]

Sloane Hawes studied how the quality of life of older animals is impacted by remaining in the care of shelters rather than being euthanized. The sample consisted of 122 dogs and 124 cats that were above the age of 7 years who were taken into a shelter in a one-year period. The article highlights the significance of preventative outreach, behaviour programs and specialized medical, and a powerful foster care system that are equipped to deal with the requirements of older animals in their homes and in a shelter. [5]

In "Ref. [6]" the authors have represented the results on predicting the pet adoption speed by using various algorithms like SVM, random forest, logistic regression.

In "Ref. [7]" the authors have discussed the results on prediction of adoption versus euthanasia among dogs and cats. This paper determines the factors on which adoption is dependent by using logistic regression.

In "Ref. [8]", the authors have represented the results of pet analytics by using their online profiles. This paper has worked on other algorithms like Gradient boosting, regression, neural networks, logistic regression etc. for speed prediction.

In "Ref. [9]", the authors have worked on pet adoption prediction using AI. In this paper, the researchers have created a website which gets updated whenever a stray animal is observed on the streets.

In "Ref. [10]", the authors have discussed the results of statistical analysis of pet adoption. In this paper, it tells us what are the preferences of people by performing ANOVA testing on the given dataset.

In "Ref. [11]", the researchers have worked on animal adoption system. In this paper, the researchers have created a website which connects the people who want to adopt and who want to give their pets virtually.

In "Ref. [12]", the authors have worked on the impact of breed identification, potential adopter perceptions and demographics, and dog behaviour on shelter dog adoptability. This paper determines what should be the characteristics of dogs for increased adoption results.

TABLE I: LITEARTURE SURVEY

Reference	Details	Algorithms used	Results	Gaps identified
[6]	Predicting the speed of adoption	Logistic regression, SVM, decision trees, random forest	Accuracy of the predictions: 0.335, 0.359, 0.321, 0.392	Accuracy was less
[7]	Predicting the speed of adoption	Catboost model	Categorises the speed into 2- within 15 days and after 15 days	Doesn't predict the exact speed
[8]	Predicting adoption speed of pets from their online profiles	Decision tree, random forest, neural networks	Accuracy of predictions: 0.58, 0.61, 0.63	Accuracy was less
[9]	AI based pet adoption prediction	Creates a website	The website will be updated whenever stray animals are observed	This only notifies whenever a stray animal is identified but doesn't predict its adoption probability
[10]	Statistical Analysis of Animal Adoption	ANOVA	Identifies the differences between breed type and outcome gender	Focuses on annual budget spent on adoption rather than predicting the speed

III. DATASET DESCRIPTION

The base dataset was taken from Kaggle which consisted of 14,993 rows and 24 columns. It consists of 3 parts:

1. It is a CSV file with detailed information including animal type (dog or cat), breed, gender, colour, fur length, sterilized, health, and description.
2. It is a JSON file of descriptions with sentiment score
3. And it is a large collection of image files.

The features which were included are: PetID, adoption speed, type, name, age, breed1, breed2, gender, color1, color2, color3, maturity size, fur length, vaccinated, dewormed, sterilized, health, quantity, fee, state, rescuer ID, video amount, photo amount, description. The adoption speed was given a numerical value based on the following:
0 - same day as listed
1 - between 1 and 7 days (1st week)
2 - between 8 and 30 days
3 - between 31 and 90 days (2nd and 3rd month)
4 - No adoption after 100 days of being listed
But this dataset was modified as shown in Table 2- the dataset regarding both cats and dogs was taken and combined together. The final dataset consists of 15 features and 1,25,326 rows which contains information about both dogs and cats. The features included are PetID, URL, type, age, gender, size, coat, breed, name, colour, posted date, contact city, contact state and adoption speed.

TABLE II: FEATURES DESCRIPTION

Feature name	Description	Domain
Pet ID	Unique ID given to each pet	String value- combination of numbers and alphabets
URL	URL of the pet image	Hyperlink
Type	Tells the category of pet	Cat, Dog
Age	Tells the age of pet	Baby, young, adult, senior
Gender	Tells the gender of pet	Male, Female
Size	Tells the size of pet	Small, Medium, Large, Extra Large
Coat	Tells how coated the pet is	Short, Medium, Long, Curly, Wire

Breed	Tells the breed of the pet	Abyssinian,Persian, American etc.
Name	Tells the name of pet	Different values of names
Colour	Tells the colour of pet	Gray, brown, black, white
Posted date	The date when the pet has been listed	Date value
Contact city	The city where the pet is available	New York, Henderson, Golden Valley etc.
Contact state	The state where the pet is available	Texas, Los Angeles etc.
Adoption speed	The speed at which pet will be adopted	0, 1, 2, 3, 4

IV. METHODOLOGY

The steps which were performed are: data collection, data pre-processing, building machine learning models, applying proper data visualisation and analysis, finding how far the models have performed well, deploying the model.

- *Data collection*: The base dataset which we are currently working on was taken from an internet source-Kaggle.com and it was modified. Two datasets were considered for this purpose- one is related to cats and the other is dogs. These two are merged and then worked upon.
- *Data pre-processing*: The data which was collected cannot be taken as it is, since it had few irrelevant values(negative number where they're not applicable), cleaning the data etc.
- *Handling missing data*: There were missing values in various features and average of the existing values was taken to fill in that data.
- *Data reduction*: In this step, the huge dataset will be reduced into a smaller one.
- *Feature Extraction*: Feature extraction is a process of renovating the fresh input data into the integral attributes and the main thing is we must preserve the data during the extraction phase. Data integrity, data confidentiality, and storage are important and retain he original information of the dataset. The raw data is processed at each stage and cleaned attributes required for the study are extracted.

The following list of analysis tasks were performed:

- *Determining if the image subject is clear*:

If cat/dog is not the first word in image keywords or description, then the subject would be marked as unclear or unfocused. Else, it would be marked as clear as shown in Fig. 1.

PetID	image.keywords	clear.subject
0008c5398	cat, small to medium sized cats, whiskers, cat like mammal, eye, khao manee, domestic short haired cat, kitten, burmilla, snout	Yes
000a290e4	dog, dog breed, dog like mammal, dog breed group, phalène, carnivoran, papillon, companion dog, moscow watchdog, puppy	Yes
000fb9572	dog, dog like mammal, mammal, vertebrate, dog breed, dog breed group, puppy, arm, snout, carnivoran	Yes
0011d7c25	cat, small to medium sized cats, cat like mammal, fauna, whiskers, domestic short haired cat, kitten, aegean cat, oriental shorthair, snout	Yes
00156db4a	dog, dog breed, snout, street dog, dog breed group, black mouth cur, dog like mammal, whiskers, rare breed dog	Yes
001a1aaad	cat, small to medium sized cats, cat like mammal, dog breed group, whiskers, carnivoran, kitten, tail, dog breed, flooring	Yes

Figure 1. Determining clear subject

- *Finding image object focus' dominant colour*:

It combines the 3 colours for a given pet and shows what is the dominant colour of a particular pet as shown in Fig. 2. It shows how many pets belong to a particular colour.

dominantColor	n
rosybrown	1079
burlywood4	933
gray20	850
gray9	687
gray19	562
salmon4	503
gray8	482
navajowhite4	435
gray32	413
gray10	398
gray21	356
gray33	342
gray61	341
gray7	334
gray60	308
gray31	295
antiquewhite3	293
peachpuff4	278

Figure 2. Finding the dominant colour

- *Finding sentiment score of description of each pet:*

As we have provided description for each pet, we determine the sentiment score by removing unnecessary words, punctuation marks, special characters and breaking them into individual words as shown in Fig. 3.

PetID	sentimentScore
0008c5398	2.00
000c21f80	1.24
000fb9572	2.33
0011d7c25	3.00
00156db4a	-1.75
001a1aaad	1.67

Figure 3. Sentiment score of description of the pet

- *Topic modelling for description feature:*

By using LDA, we determine the hidden features for the description of each pet and we rename the columns as doc_topic 1, doc_topic 2, doc_topic 3, doc_topic 4, doc_topic 5 as shown in Fig. 4.

PetID	doc_topic_1	doc_topic_2	doc_topic_3	doc_topic_4	doc_topic_5
58af165cd	0.0002407595	0.0002407106	0.9839596032	0.0002407647	0.0153181620
86547ad6d	0.4146193076	0.1680716461	0.0002455274	0.3868039009	0.0302596180
a63f6c6a1	0.5514827925	0.1102640706	0.0003533545	0.3070933243	0.0308064581
3e2cc2690	0.8411213943	0.0230723699	0.0005373701	0.0005376459	0.1347312198
427acc248	0.8222218310	0.1750551684	0.0009074380	0.0009079141	0.0009076485
9967c6cb6	0.5828149735	0.0576174315	0.1743712229	0.0003639845	0.1848323875

Figure 4. Topic modelling for description

- *Topic modelling for image keywords feature*:

The same is performed for the image keywords and the columns are renamed as keyword_topic 1, keyword_topic 2, keyword_topic 3, keyword_topic 4, keyword_topic 5 as shown in Fig. 5.

PetID	keyword_topic_1	keyword_topic_2	keyword_topic_3	keyword_topic_4	keyword_topic_5
ac3f4aec5	0.002318314	0.002318314	0.5703341	0.002318314	0.422710936
03c1e47b4	0.002559224	0.056582884	0.9357394	0.002559224	0.002559224
0605aab21	0.002559224	0.002559224	0.7182754	0.002559224	0.274046960
083687ba1	0.002559224	0.002559224	0.9897631	0.002559224	0.002559224
0a0862717	0.002214102	0.002214102	0.3535033	0.002214102	0.639854418
1b91e1082	0.002559224	0.002559224	0.9897631	0.002559224	0.002559224

Figure 5. Topic modelling for image keywords

- Replacing the name column with 2 values- Has Name, No Name and add WordCount column:

If a particular animal has name, we assign it as Has Name and if it doesn't, we assign it as No Name. In this step, we also add the column of WordCount- count the number of words in the description for each pet as shown in Fig. 6.

PetID	Type	Name
0008c5398	2	Has Name
000a290e4	1	Has Name
000c21f80	2	Has Name
000fb9572	1	Has Name
0011d7c25	2	Has Name
00156db4a	1	Has Name

keyword_topic_1	keyword_topic_2	keyword_topic_3	keyword_topic_4	keyword_topic_5	Word.Count
0.577677860	0.413754111	0.002856010	0.002856010	0.002856010	47
0.003457432	0.003457432	0.986170272	0.003457432	0.003457432	26
0.002699484	0.877476462	0.002699484	0.114425086	0.002699484	211
0.003718444	0.003718444	0.903578602	0.085266068	0.003718444	27
0.989202064	0.002699484	0.002699484	0.002699484	0.002699484	8
0.002856010	0.067428454	0.924003517	0.002856010	0.002856010	75

Figure 6. Adding Wordcount column and converting name to Has name and No name

- One hot encoding and factoring for categorical features:

We convert the features with categorical values like dewormed, sterilized, quantity, size, photoamt etc. into numerical values by using one hot encoding as shown in Fig. 7.

PetID	AdoptionSpeed	Type	Name	Age.Quantile	Gender	MaturitySize	FurLength	Vaccinated
0008c5398	3	2	1	4	1	2	2	1
000a290e4	2	1	1	1	2	2	2	2
000c21f80	NA	2	1	2	1	1	1	2
000fb9572	3	1	1	1	2	2	1	2
0011d7c25	2	2	1	1	1	2	1	2
00156db4a	1	1	1	1	1	2	1	2

Figure 7. One hot encoding

- Building a confusion matrix for actual and predicted classes:

Finally, we build a confusion matrix for the model if it has predicted correctly or not as shown in Fig. 8.

```
=========================================================================
Confusion Matrix: Row labels: Actual class; Column labels: Predicted class
       0    1    2    3    4  Error            Rate
0      1   16  259    3   37 0.9968 =      315 / 316
1      0 2158  336    0  106 0.1700 =    442 / 2,600
2      0  218 2607  229  414 0.2483 =    861 / 3,468
3      0    3  121 2587  103 0.0807 =    227 / 2,814
4      0   63   23   18 3372 0.0299 =    104 / 3,476
Totals 1 2458 3346 2837 4032 0.1538 = 1,949 / 12,674
```

Figure 8. Confusion matrix

A. Algorithms Used:

Classification: This is a task where a particular value will be categorised into pre-defined classes. It is a process of creating classes that represent users and use cases. Class Probability Estimation tries to predict how to classify each single individual data asset. Based on the specific question to be answered, classes are created. Other qualities of the asset are evaluated in making the prediction.

Factoring: This is a technique to convert categorical values into numerical values. For example, considering the age feature- we have values like adult, young, old etc. They will be converted into 0, 1 and 2 respectively. This is called one-hot encoding.

Regression: The most commonly-used forecasting method is the Regression method. Regression can be confused with classification methods because to predict an outcome, the process of using known values is the same. It is a model used to predict continuous value. Linear regression, Logistic Regression, and Polynomial Regression are some types of regression.

V. RESULTS

TABLE III. RESULTS

Analysis task	**Result obtained**
Determining if the image subject is clear	Yes/No according to the first word of Image keyword
Finding image object focus' dominant colour	Number of pets belonging to a particular colour
Finding sentiment score of description of each pet	Positive/Negative value
Topic modelling for description feature	Finds the hidden features for description and adds the columns as doc_topic 1 to 5.
Topic modelling for image keywords feature	Finds the hidden features for image keywords and adds the columns as keyword_topic 1 to 5.
Replacing the name column with numerical values and adding WordCount column	Has Name or No Name and sum of no. of words in description.
One hot encoding and factoring	Converts values like small, medium, large, young, old, baby etc. into numerical values.
Building a confusion matrix	For predicted and actual values.

VI. Conclusions And Future Scope

The findings of this paper highlight the features which are best chosen when it comes to adopting a pet. We determine the speed of the adoption of the pets as 0, 1, 2, 3 and 4 based on how many days after the pet has been posted, it has been adopted. We use both image data and text data like description and image keywords for the pets to determine this. It is clear that features like colour, age and breed play a vital role when a pet is being adopted.
The task for this project is to predict pet adoption speed given related information such as biological details, descriptions. We use the machine learning algorithm H2O on the given data which has given following observations:
MSE: 0.321
RMSE: 0.567
Logloss: 0.8567
Null Deviance: 37022.39
Residual Deviance: 21717.15.
Based on the result, visual information such as photos and videos of the pet seems to have the most influence on the adoption decision.
For future work, we plan to experiment with more machine learning models, and try different ensemble methods in the hope of an increase in accuracy. Given the importance of visuals and the available image data, we would like to take full advantage of the dataset and incorporate these images into our deep learning models using computer vision algorithms. Different model structures and hyperparameter combinations can be tested and adjusted accordingly if more time and computational resource are available.

References

[1] Bartlett, P. C., Bartlett, A., Walshaw, S., & Halstead, S. (2005). Rates of euthanasia and adoption for dogs and cats in Michigan animal shelters .Journal of Applied Animal Welfare Science, 8(2), 97-104
[2] Brown, W. P., & Morgan, K. T. (2015). Age, Breed Designation, Coat Color, and Coat Pattern Influenced the Length of Stay of Cats at a NoKill Shelter. Journal of Applied Animal Welfare Science, 18(2), 169- 180
[3] Rachael Mozes, David L Pearl, Lee Niel, J Scott Weese(2017) Epidemiological investigation of euthanasia in an Ontario animal shelter, Journal of Feline Medicine and Surgery , 20 (6), 479-486
[4] Lord, E., Widmar, N. O., & Litster, A. (2014). Economic impacts of adoption and fundraising strategies in animal shelters. Preventive Veterinary Medicine, 113, 423–429
[5] Sloane Hawes, Josephine Kerrigan, and Kevin Morris (2018) Factors Informing Outcomes for Older Cats and Dogs in Animal Shelters, Animals , 8(3), 36.
[6] Zhang K, Zhang S. PetFinder challenge: Predicting pet adoption speed. Accessed: Nov. 2021;21.
[7] Lepper M, Kass PH, Hart LA. Prediction of adoption versus euthanasia among dogs and cats in a California animal shelter. Journal of Applied Animal Welfare Science. 2002 Jan 1;5(1):29-42. [8] Zadeh A, Combs K, Burkey B, Dop J, Duffy K, Nosoudi N. Pet analytics: Predicting adoption speed of pets from their online profiles. Expert Systems with Applications. 2022 Oct 15;204:117596.
[8] Kajbaje S, Sawant R, Patil RL. AI-Based Pet Adoption System.
[9] Anand R, Cherian SM, Ravi S, Agarwal V, Sampath A, Shreenidhi S, Menon A, Srinath KR. Statistical Analysis of Animal Adoption using R.
[10] Pavao-Zuckerman B, Reitz EJ. Introduction and adoption of animals from Europe. Handbook of North American Indians. 2006;3:485-91.
[11] Bradshaw YF. The Impact of Breed Identification, Potential Adopter Perceptions and Demographics, and Dog Behavior on Shelter Dog Adoptability (Doctoral dissertation, The Ohio State University).