3244-2010-0014 - The Pawbability of Adogtion

Authors: Chua Xin Xuan, Lim Xi Chen Terry, Lim York Tee Gorden, Luk Chi Heng, Tan Jie Yi, Tan Yoong Kang Colin
Email: e0425690@u.nus.edu, e0406494@u.nus.edu, e0309693@u.nus.edu, e0310461@u.nus.edu, e0426306@u.nus.edu, e0325403@u.nus.edu

Abstract

In this report, we aim to predict the adoptability of pets in animal shelters given its portfolio on a web profile page. Additionally, we want to identify which traits & profile pictures are desirable when pet adopters choose to adopt a pet. This will be done via the use of Decision Trees and a Convolutional Neural Network models to deal with the complex mix of variable types in our data. Based on our prediction models, we will examine the significant variables and derive insights on the possible actions following our results.

1 Introduction

Some animals seem to be able to get adopted easily into their forever homes, while others only have bleak futures ahead of them as they continue to remain in the shelter.

With the growing popularity of online adoption drives[1], we wish to build a prediction model to predict if an animal would get adopted based on its online portfolio.

Furthermore, to understand this phenomenon better, we also identify which attributes in a pet's web profile, such as gender or age, play a more significant role in determining whether it gets adopted or not. We will also be looking into the profile pictures of the online portfolios to see if there is any correlation between profile pictures of animals that do or do not get adopted.

For this report, we will make use of Decision Tree and Convolutional Neural Network models to try to provide an accurate prediction as well as interpret the results to discern and understand the significant attributes and their behaviour.

Through this, we hope to improve the chances of these animals being adopted, which will help increase the shelters' ability to take in more stray animals, decrease the number of animals being euthanized in shelters or come up with preemptive solutions to cater to animals identified to have trouble in getting adopted.

This report will explain the machine learning methods and decisions we have employed, our subsequent predictions and how we interpreted the findings. For each of these methods, we will first elaborate on its methodology, then evaluate how accurate the predictions generated are based on both real life observations as well as through preliminary exploratory data analysis.

2 Related Work

The dataset we will be working with comes from Petfinder, a non-profit organization. The dataset used in particular comes from the Kaggle Competition[2] hosted by Petfinder Malaysia in 2018 with the intention of finding the speed in which an animal gets adopted. For the purposes of this report, we only considered animals who do or do not get adopted.

There already exists several studies and reports on the dataset that also choose to focus on both building models to determine a pet's adoption as well as identifying the significant attributes. Such reports include Zhang's Predicting Adoption Speed for Petfinder[3], and the various public submissions from the Kaggle Competition.

However, most of these analyses chose to utilise a full neural network deep learning model. Such examples include the best performing Kaggle submission by the team Wodori[4], prioritising accuracy but leaving little to none interpretability of the results due to the 'black-box' nature of neural network models.

Such works also fail to conduct further testing to determine the significance or understand the correlations between adopted and unadopted pet profiles.

Others, like Zhang, instead choose to only focus solely on the tabular data by modeling them after classical machine learning algorithms, which provides useful insights and interpretation of the data. However, they leave out attributes which we consider are potential key factors that impact the decision behind the adoption of a pet (such as the pictures uploaded on the adoption website).

Taking the above shortcomings into mind, we have decided to find the halfway point between the two extreme directions others have taken. We hope to achieve this by employing both highly interpretable machine learning algorithms as well as deep learning neural networks with subsequent testing. We will be able to provide not only accurate results, but a more holistic and complete interpretation of the findings using the data we have available to us, as compared to prior works.

3. Main Methodology

Our dataset comprises a plethora of variable types, notably photos and tabular data regarding characteristics of the pet for adoption.
As such, our initial plan was to build a multi-input model, where we feed the tabular data into a Dense network and image data into a CNN network, then merge them and pass it into a common dense layer.

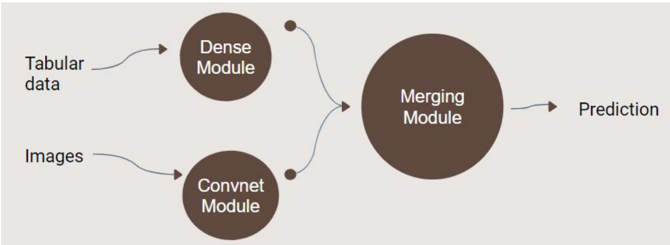


Fig 1: Multi-input model displaying insertion of both tabular data and images into their respective connections into a merging model to generate a prediction

However, due to the 'black-box' nature of deep learning neural networks, we will not be able to properly or accurately discern what are the 'significant' variables and their level of importance using such an approach.

[1] Lock, C, (2020), *Covid-19 heroes: Animals find fur-ever homes with online adoption drives,* https://www.straitstimes.com/lifestyle/animals-find-fur-ever-homes-with-online-adoption-drives
[2] https://www.kaggle.com/c/petfinder-adoption-prediction/overview
[3] Zhang, E.Z, (2020),Predicting Adoption Speed for Petfinder, https://towardsdatascience.com/predicting-adoption-speed-for-petfinder-bb4d5befb78c
[4] https://www.kaggle.com/naka2ka/stack-480-speedup-groupkfold-with-no-dict

Hence, we decided to have two separate approaches:

Approach 1: Decision Trees (DT)

Decision Trees will be used to track and model the mix of categorical, continuous and ordinal features. Due to how we had cleaned and transformed the data and the presence of different types of variables, we have tried implementing different classification methods as well, which are also included in our code.

These models include Multinomial Naive Bayes, Logistic Regression and K-Nearest Neighbors. We would use them as our baseline models (by comparing their workability, accuracy and interpretation) and compare them in our subsequent final model for tabular data during our evaluation.

Ultimately, we chose to use Decision Trees as not only does it give a high enough accuracy score, we are able to easily interpret and determine important variables by simply observing the nodes and branches of the tree. To corroborate our decision, we performed Decision Trees on our overall dataset to verify if it is implementable, feasible and effective, before doing our split analysis on cats and dogs.

We had considered using Random Forest as well which would produce a better accuracy score at the cost of interpretability. However, as we are much more interested in the significance of our variables, we valued the high interpretability of the decision tree over a better prediction score.

Approach 2: Computer Vision (CV)

Computer Vision was used to handle the important spatial informational aspects in the images of animals in our data set. We believe that a Convolution Neural Network is a good choice due to its high accuracy in image classification. Images have high dimensionality, thus CNN will serve to reduce the number of parameters required, while preserving important features. This can decrease the problem of overfitting which will lead to high test cost. Hence, resulting in a high accuracy for image classification.

We had also converted the AdoptionSpeed column to an Adoption column that contains only binary values. Adoption speed of 0 - 3 is classified as Adopted, while adoption speed of 4 is classified as not Adopted.

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

Fig 2: Table documenting our relabelling from the original dataset (Adoption Speed) to simply Adoption

4 Decision Trees

4.1 Data Manipulation

To deal with categorical, ordinal and continuous data, we want to manipulate our inputs into a categorical or ordinal format to aid the decision making in our decision tree. Hence, we converted some continuous variables to ordinal data based on their respective quantiles. Columns “Breed1” and “Breed2” have high dimensions in which they can take around 300 different values. Rather than having over 300 more variables from one-hot encoding, we reduced the dimensions by collapsing this into a single binary variable: 1 indicates the animal is a pure breed while 0 indicates its a mixed breed. Additionally, we dealt with “Age” specially, as the difference in months at earlier stages bear more significance than during old age. Thus, a more intricate interval is used to bin “Age” into ordinal data, where we split between [0-1], (1-2], (2-3], (3-6], (6-12], and (12 and above), to account for younger animals in greater detail.

The data set provided only consists of cats and dogs, and the distribution is shown as below:

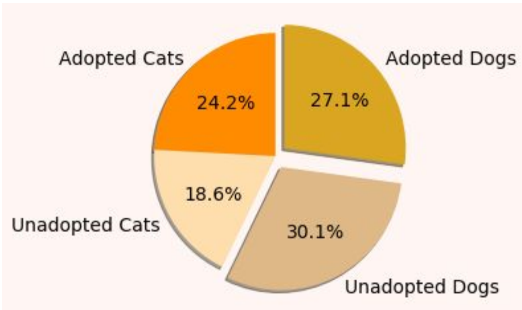


Fig 3: Breakdown of class distribution in our revised dataset

4.2 Decision Trees Modelling

After cleaning the data, two separate Decision Trees were constructed, with prior information that the important features for cats and dogs may differ. We then create a general flow in constructing each of these trees. Firstly, we tried two criterions: Gini Index and Entropy. Secondly, we iterated and plotted accuracy scores across an interval of lengths for maximum tree depth (max_depth). Thirdly, we had identified the best max_depth and reviewed accuracy scores near that depth for pruning. Lastly, we finalized the choice of max_depth and criterion, then implemented the Decision Tree for the dataset. With the tree constructed, we can output Accuracy Scores and compare across [Accuracy, Specificity, Sensitivity, F1-Score and AUC], where we tuned it with a validation set and used a test set to showcase final scores.

Once we obtain a decision of suitable accuracy, we can identify the important attributes in the decision making process. This is done via “model.feature_importances_”, which is based on Gini importance, where the amount of “impurity” this feature reduces is computed and normalized. The greater the reduction, the better it is.

4.2.1 Decision Trees for Cats

Having prior beliefs that cats and dogs may have different important attributes, we split our analysis and created Decision Trees for each type. Following our modelling above, we chose the criterion to be ‘Entropy’ and set the ‘Max_depth’ at 6. This is because the optimal accuracy score was achieved at max_depth = 6. We went with it as its score is significantly larger than other max_depth values and has enough representation power while being able to specialize and determine more important features.

4.2.2 Evaluation for Cats (DT)

Evaluation Metric	Score (To 3 significant figures)
Accuracy Score	0.632
Specificity	0.566
Sensitivity	0.692
F1-Score	0.664
AUC	0.629

Fig 4: Evaluation Metrics with the Cat Dataset

4.2.3 Findings for Cats (DT)

For cats, the important features (in descending order): Age, Amount of Photos, Fur Length, Sterilized and Maturity Size. Age is the most pivotal feature with the largest Gini importance of 0.579 compared to PhotoAmt at 0.152. Notably, the relationship between MaturitySize and the adoptability of cats is in contrast with dogs, which will be specified later.

Maturity Size for Cats:
(1 = Small, 2 = Medium, 3 = Large, 4 = Extra Large)

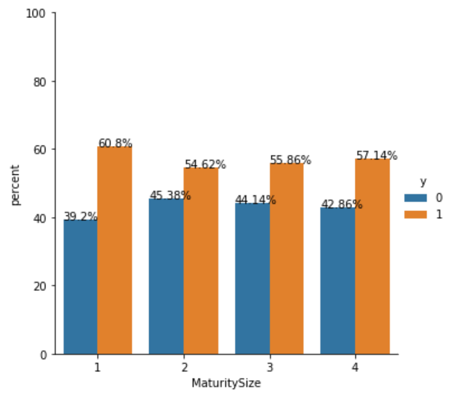


Fig 5: Barchart of the proportion of adopted and unadopted cats against their Maturity Size. *Blue bar depicts "not adopted", orange depicts "adopted".*

Cats of the smallest size are heavily favoured, but sizes larger than small have similar chances of being adopted.

4.2.4 Decision Trees for Dogs

Similarly for dogs, we chose the criterion to be 'Entropy' and set the 'Max_depth' to be 8. This was because the optimal accuracy score was achieved at 'Max_depth' = 8. We went with it as once again, its score is significantly larger than other max_depth values and has enough representation power while being able to specialize and determine more important features.

4.2.5 Evaluation for Dogs (DT)

Evaluation Metric	Score (To 3 significant figures)
Accuracy Score	0.608
Specificity	0.659
Sensitivity	0.552
F1-Score	0.574
AUC	0.606

Fig 6: Evaluation Metrics with the Dog Dataset

4.2.6 Findings for Dogs (DT)

For dogs, the important features (in descending order): Age, Amount of Photos, Maturity Size, Fur Length and Sterilized. Despite having the same attributes as cats, the importance across these features are rather equal, having similar values for Gini importance where Age is only 0.07 higher than Amount of Photos. Previously mentioned, the relationship for Maturity Size and adoptability of dogs differs in distribution from cats, hence we will showcase the contrast.

Maturity Size for Dogs:
(1 = Small, 2 = Medium, 3 = Large, 4 = Extra Large)

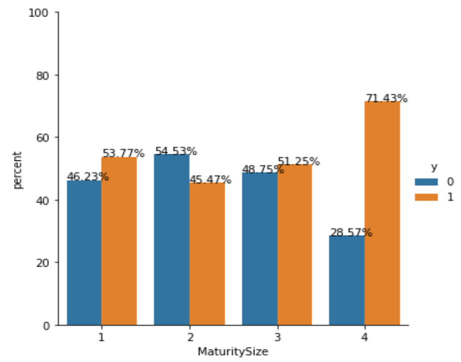


Fig 7: Bar chart of the proportion of adopted and unadopted dogs against their maturity size.

Dogs of the largest size are heavily favoured, but sizes larger than that have similar adoption chances of being adopted.

4.3 Overall Discussion

With both our Decision Trees pruned and finalized, we can now evaluate our model across the different baseline models mentioned in Section 3. We can compare the accuracy scores below:

Classifier	Score
Multinomial Naive Bayes	0.602
K-Nearest Neighbors	0.518
Logistic Regression	0.628
Decision Forest	0.639

Decision Tree	Score
Cats	0.632
Dogs	0.608

Fig 8: Accuracy Scores using Baseline Models

Our two Decision Trees performed generally well in terms of prediction, despite being slightly lower in scores than some baseline models. However, we note that these baseline models are difficult to interpret which variables bear more weight, despite being fast in implementation. Therefore, since our decision trees have similar performance in terms of accuracy and are able to determine feature importance easily, it was optimal to use Decision Trees as our final model; even though it may require more meticulous details in structuring, where pruning and some dimensionality reduction was conducted.

After evaluating, we would like to consolidate our findings from both Decision Trees. Besides Maturity Size, all other important attributes were common in both Cats and Dogs. We note that they vary in the order of importance, where some attributes bear more weight depending on the type of animal. We analyzed their distributions separately for each of these features and realized that they are very identical. Hence, a combined evaluation is depicted below, to reduce repetition.

Age:
As the pet grows older, its chances of adoption severely decreases, especially in the earlier months.

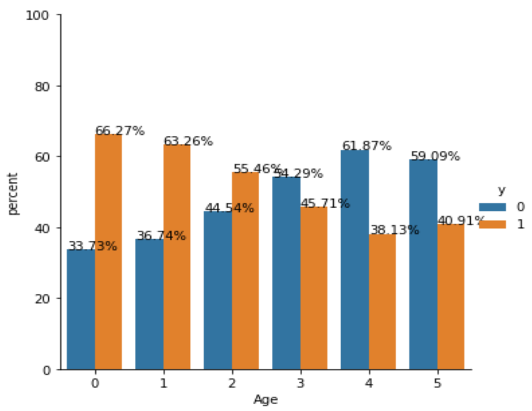


Fig 9: Bar chart of the proportion of adopted and unadopted dogs against their age.

Fur Length:
Longer Fur Length seems to be more appealing, as seen by the percentage of adoption increasing as FurLength increases. This could imply that fluffy and furry pets are more attractive as compared to bare-skinned ones.

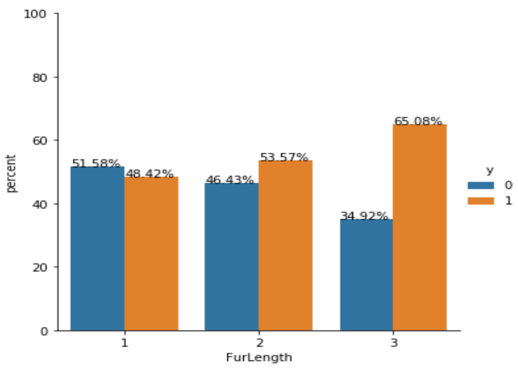


Fig 10: Bar chart of the proportion of adopted and unadopted dogs against their Fur Length

Sterilized:
(Binary, 1 = Has been sterilized, 0 = Has not been sterilized):
Surprisingly, pets who are sterilized have a much lower chance of being adopted.

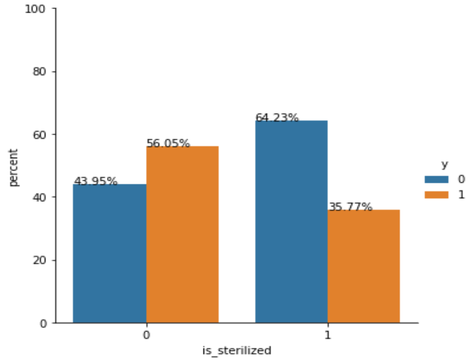


Fig 11: Bar chart of the proportion of adopted and unadopted dogs against if they have been sterilized

Amount of Photos:
Surprisingly, pets with a smaller amount of photos tend to be adopted more often.

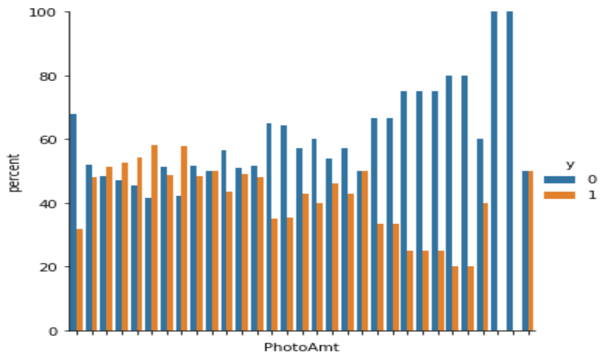


Fig 12: Bar chart of the proportion of adopted and unadopted dogs against their number of photos.

However, we note that when pets have no photos at all, their chances of adoption are much lower, emphasizing the importance of our Computer Vision model to derive more significant insights.

Concluding our segment on dealing with tabular data through Decision Trees, our next section would elaborate on our Computer Vision model in detail, in order to determine the significant factors in the images of the pets.

5 Computer Vision

5.1 Method

Data exploration and analysis was first carried out to prepare the data for training.

1. Rows containing no images were dropped, because the model requires images as data instead of tabular data.

2. The first image for pets which have more than one photo was used for the model to ensure fairness when comparing between each pet. Furthermore, since the first image is the profile picture of the pet in the Adoption webpage, it will most likely have the greatest impact on the potential pet adopters when they first see an animal.

3. For the percentage of data to be used for training, testing and validation, 80% of the data was first used for training and 20% for validation. However, we realised that we were unable to use the Kaggle platform for testing as the problem was simplified from a multi-class problem to a binary-class problem. Thus, the bottom

10% of the data was used for testing. Out of the remaining 90%, 80% was used as training and 20% for validation.

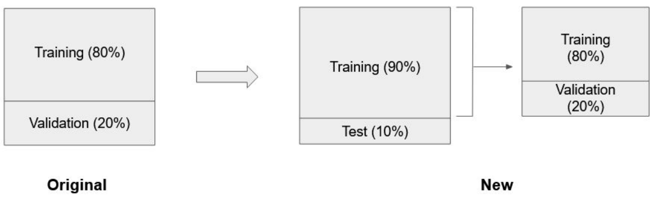


Fig 13: Train Test Split of Dataset

4. Similar to decision trees, we have also split the data into cats and dogs, because we believe that potential pet adopters may be looking out for different features for different pets. For example, dog adopters may be looking out for dogs which are bigger in size and darker coated, while for cats, they may be looking out for cats which are smaller in size and lighter-coated.

5.2 Model

The model below is used for both cats and dogs.
Model: Pre-trained Xception model + Additional hidden layers
Activation function: Sigmoid
Optimizer: Stochastic Gradient Descent
Loss function: Binary Crossentropy
Metric: Accuracy

```
Model: "sequential"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
xception (Functional)        (None, 8, 8, 2048)        20861480
_________________________________________________________________
global_average_pooling2d (Gl (None, 2048)              0
_________________________________________________________________
output (Dense)               (None, 1)                 2049
=================================================================
Total params: 20,863,529
Trainable params: 2,049
Non-trainable params: 20,861,480
_________________________________________________________________
```

Fig 14: Summary of Model

5.3 Evaluation

Model	Accuracy of Dog Model (To 3 significant figures)	Accuracy of Cat Model (To 3 significant figures)
Simple CNN Model	70.2%	72.5%
DenseNet121	65.0%	73.1%
InceptionV3 Model	70.5%	73.1%
Xception Model (w/o Data Augmentation)	72.0%	73.0%
Xception Model (w Data Augmentation)	72.5%	73.5%

Fig 15: Accuracy Scores of Models

Our baseline model was a simple image classification model built from scratch. However, to further improve the accuracy of such a model would take a longer period of time and more resources to fine tune it. Hence, we decided to utilise transfer learning[5] which is to use a pre-trained model[6] with its trained weights as the base model. A pre-trained model has a good representation of features for over a million images belonging to a thousand different categories. It can act as a good feature extractor for new images making it suitable for computer vision problems.

Among all the various pre-trained models tested, we ultimately chose to use Xception as it has the highest accuracy. Furthermore, it has been previously trained on images for image classification tasks which makes it a good fit for our task.

Furthermore, our model requires a probability value as an output to decide whether a pet is adopted or not. Thus, we have narrowed down to either Softmax or Sigmoid among all the various activation functions. Comparing between these two, Softmax is used for multiclass problems, while Sigmoid is used for 2-class problems. Hence, sigmoid is the most suitable activation function to be used in our model.

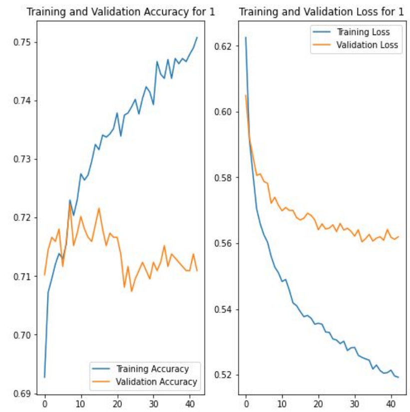


Fig 16: Graph of Training and Validation Accuracy & Loss for Xception Model (w/o Data Augmentation)

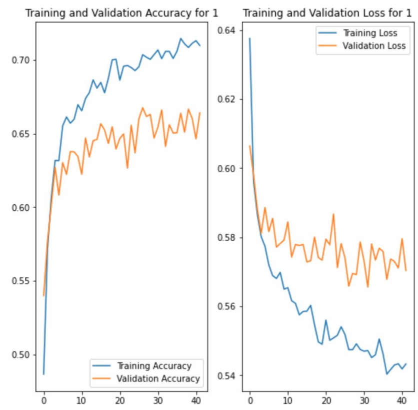


Fig 17: Graph of Training and Validation Accuracy & Loss for Xception Model (w Data Augmentation)

After running our model and plotting the trained loss and validation loss for each epoch, we observed that there was a large difference between the 2 losses due to overfitting as shown in Fig 16. This

[5] Dipanjan(DJ) Sarkar, (2018), *A Comprehensive Hands-on Guide to Transfer Learning with Real-World Applications in Deep Learning*, https://towardsdatascience.com/a-comprehensive-hands-on-guide-to-transfer-learning-with-real-world-applications-in-deep-learning-212bf3b2f27a

[6] Billy Lamberta, Bastian Eichenberger and others, (2017), *Transfer learning and fine-tuning*, https://www.tensorflow.org/tutorials/images/transfer_learning

might be because the amount of data which we can work with has shrunk significantly after filtering and splitting the data into dogs and cats. With a small dataset for the model, the model kept seeing the same instance over time across each epoch, resulting in overfitting. Hence, we have used image augmentation whereby we applied some image transformation operations to the dataset (e.g. rotation, shearing, translation, zooming, etc.) to produce new, altered versions of existing images. This would artificially increase the size of our training data while maintaining the integrity of the data. Thus, as seen in Fig 17, the difference between the trained and validation loss were reduced when image augmentation was used.

5.4 Discussion

Possible reasons for pets not being adopted include:

1. Dilution of images with text and other images
2. Collaged images
3. Background noise (e.g. presence of human)
4. Irrelevant images
5. More than one pet in the photo

Some examples are as shown below:

Fig 18A: Image with a lot of dilution of other irrelevant images and text

Fig 18B: Image with a girl present (Background Noise)

The profile photo which is the first image of every pet is crucial as it is the photo that gives potential adopters their first impression. This affects whether a potential adopter would click in to find out more about the pet.

To help boost a pet's chances of getting adopted, we believe that the pet should be the main focus in its profile photo. It would also yield higher adoptability for the pet if it were to look directly at the camera.

6 Conclusion

6.1 Summary

Although both models have an accuracy score of approximately 60% to 75%, this was slightly below our expectations. Our research is still useful as our findings have suggested that pet adopters place a high importance on the quality of photos and age of the animal. Higher quality photos and younger pets tend to yield a much higher chance of being adopted. We also found that furry animals are generally more well liked and maturity size plays a vital role. Therefore, in order to improve the adoptability of a pet, it is more effective to work on these key details, disregarding other noisy attributes that may seem important but do not actually play a large role. This would save time, costs and resources, ultimately improving the overall adoption rate.

6.2 Shortcomings

We understand that many other important external factors such as the number of volunteers and geographical location of the shelters were not taken into account in our project due to the lack of data.

Our decision tree could have used more extensive pruning with regards to other hyperparameters[7] such as min_impurity_decrease, min_impurity_split, min_samples_leaf, max_leaf_nodes and max_features as opposed to just pruning based on max depth.

Furthermore, if given enough time, bounding boxes could have been implemented in the images to pinpoint the position of the animal in the image to allow for better identification of the animal.

A possible ethical issue that may arise is the possibility that sheltered animals may receive an unequal level of treatment based on the model's findings. Shelters may choose to allocate more resources on animal's predicted by the model to not be adopted. In the event that the prediction is misclassified, animals already struggling to get adopted may face a tougher time.

6.3 Future Work

Overall, with all the important features identified, we hope that pet shelters could make use of our findings to make guided and focused improvements of the profile of the animals in their shelters, thereby increasing the chances of these animals finding their forever homes.

We wish to follow up and disseminate our research and findings to animal protection organisations in Singapore such as SPCA, SOSD and ACRES. We hope that we can apply such methods of Machine Learning and research into Singapore's context to better understand the local landscape of the pet adoption culture.

[7] Sodner Yildirim, (2020), *Hyperparameters of Decision Trees Explained with Visualizations*, https://towardsdatascience.com/hyperparameters-of-decision-trees-explained-with-visualizations-1a6ef2f67edf

References

[1] Lock, C, (2020), *Covid-19 heroes: Animals find fur-ever homes with online adoption drives*, https://www.straitstimes.com/lifestyle/animals-find-fur-ever-homes-with-online-adoption-drives

[2] Kaggle, *PetFinder.my Adoption Prediction*, https://www.kaggle.com/c/petfinder-adoption-prediction/overview

[3] Zhang, E.Z, (2020), *Predicting Adoption Speed for Petfinder*, https://towardsdatascience.com/predicting-adoption-speed-for-petfinder-bb4d5befb78c

[4][kaggler-ja] Wodori,(2019), *stack 480 + speedup + GroupKFold with no dict,* https://www.kaggle.com/naka2ka/stack-480-speedup-groupkfold-with-no-dict

[5] Dipanjan(DJ) Sarkar, (2018), *A Comprehensive Hands-on Guide to Transfer Learning with Real-World Applications in Deep Learning*,https://towardsdatascience.com/a-comprehensive-hands-on-guide-to-transfer-learning-with-real-world-applications-in-deep-learning-212bf3b2f27a

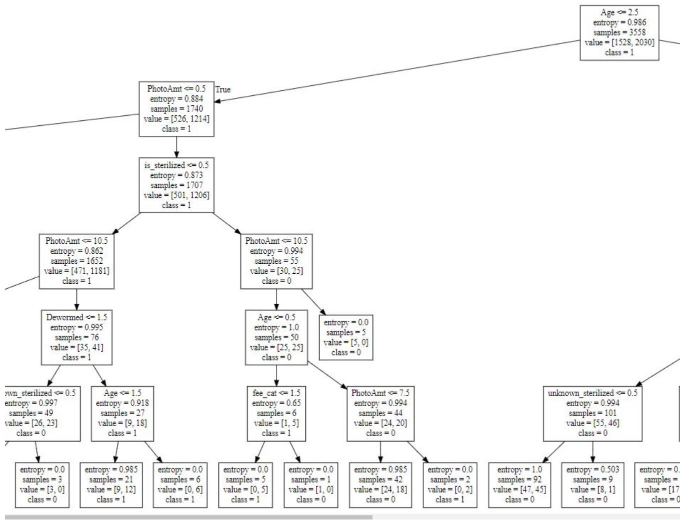
[6]Billy Lamberta, Bastian Eichenberger and others, (2017), *Transfer learning and fine-tuning*, https://www.tensorflow.org/tutorials/images/transfer_learning

[7]Sodner Yildirim, (2020), *Hyperparameters of Decision Trees Explained with Visualizations,* https://towardsdatascience.com/hyperparameters-of-decision-trees-explained-with-visualizations-1a6ef2f67edf

Appendix

Link to our project on github:
https://github.com/CS3244-2010-0014/Pawbability-of-adogtion

Portion of Decision Tree for Cats:



Regarding both full Decision Trees, they can be accessed in our github under cat.dot (for Decision Trees on Cats) and dog.dot (for Decision Trees on Dogs). We have chosen to put the full tree in the appendix as it is very large, and we feel that we have captured the important information we require in our report, acknowledging that our main focus was to identify significant attributes in determining the probability of adoption. To view these dot files, input the content into http://webgraphviz.com/ to view the full Decision Trees.