U.PORTO
FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

PetFinder: Pawpularity score

Leonor Moura up201906437, Maria Amaro up202008463, Margarida Nogueira up202204113

Abstract — **This project is based on an idea from a Kaggle competition. Its objective is to predict user's engagement with a pet's profile in an adoption website based on the photo associated, allowing the selection of more compelling images. The final goal is to assign each photo a Pawpularity score. For this purpose, we will compare three different approaches for obtaining the value of the target: 1) using just the Metadata; 2) using just the images' features; 3) combining features from images and tabular data. Ultimately, we intend to investigate novel ways to achieve a popularity score for the animal's photos, using Machine Learning Techniques.**

I. INTRODUCTION

In Portugal, around 80 thousand animals per year are living in shelters, hoping to be given a loving home, which underlines the importance of adoption. (Pires, Agosto 2023) Thus, the goal of this project is to be able to predict engagement with a pet's profile in an adoption website based on its photograph, since it has been proven that the profile pictures are extremely important in the outcome of adoption. Although being an anecdotal resource, 65% of the adopters viewed online photos of dogs before adopting them. (How Photos Are Important To Pet Adoption: A Study, 2018) In fact, the increase in Internet usage has changed how adoptable pets are presented to the public having been demonstrated the strong association between high-quality images and several photo traits (nonblurry photos, eye contact, the animal not being small and the angle at which the photos were taken) and the time the animal remained on the shelter before adoption. (Witte, 2014) Recent studies also suggest that cute images stimulate the pleasure centers of the brain which are closely related with positive emotions of human beings. (Melanie L Glocker 1, 2009) With this purpose, we attempt to develop a computational model capable of quantifying the appeal of an animal photo, considered as "Pawpularity Score". This score was originally calculated by the Petfinder website (PetFinder.my, s.d.) and derived by each pet profile statistics at the listing pages. So, through the utilization of machine learning techniques, we aim to learn and identify key features that contribute to perceived attractiveness, allowing us to objectively evaluate and rank the appeal of pet images. In fact, in computer vision, there are many works aiming to recognize ordinary expressions, such as happiness and sadness. (Hiroshi Nittono, 2012) However, cuteness is beyond these traditional expressions and has higher-level semantics, making it much more complicated to identify. We hope to objectify regularities in all these important aspects, essentially cuteness and image quality, allowing them to be analyzed and learned (Kang Wang, 2015), since this innovative application of technology is promising in providing a manner to positively impact animal welfare and adoption rates.

The project concept and the dataset used are based on a competition from Kaggle. (PetFinder.my - Pawpularity Contest, 2021) This work expands the concept of the competition by comparing three different methods for obtaining the value of the target. Firstly, using the metadata, secondly using just the images' features, and finally, combining the features from the previous approaches.

II. METHODOLOGY

A. Dataset

The dataset is composed by 9624 observations containing image and tabular data. The dataset was firstly divided into a training set (for training and validation) and a test set (for evaluation) with a split of 80%-20%.
The images are sourced from the Petfinder Website in a '.jpg' RGB format and the tabular data contains binary features of elements in the image and the calculated value of "Pawpularity". Each observation is linked to a unique Pet Profile ID. This target value was obtained on the website by using an algorithm that normalizes the traffic data across different pages, platforms (web & mobile) and various metrics. The dataset contains metadata for all the images and presents 12 features labeled with 1 (yes) or 0 (no) which include: focus, eyes, face, near, action, accessory, group, collage, human, occlusion, info and blur. These features are not accounted for on the original algorithm for Pawpularity score but we have included them in our analysis of the data due to the potential benefit to co-relating them to a photo's attractiveness. In order to achieve more consistent results throughout the model development process, we have considered as target variable a normalization of the Pawpularity score from 0-100 into the range 0-1.

B. Distribution of data

The distribution of the target variable on the dataset was analyzed with the help of a distribution plot. When analyzing this distribution, it was noticeable that the target variable followed a shape close to a Gaussian distribution but there were some outliers identified. The dataset presented several values of images scored with 1.0, which disrupted the statistical trend. Consequently, we tested the hypothesis that these values were outliers due to an error of the algorithm or misclassification and would not be helpful in predicting the actual true value of the target variable. When constructing the models, they were trained with the full training dataset and with a dataset with these values of 1.0 removed (outlier free dataset). The distribution of the target variable on both datasets presents the statistics shown in Table 1 and the distribution plots can be compared in Figure 1.

Table 1 – Statistics for the distribution of the target variable on both datasets.

	TRAINING SET WITH OUTLIERS	**TRAINING SET WITHOUT OUTLIERS**
COUNT	6343	6181
MEAN	0.3788	0.3626
STD	0.2027	0.1783
MIN	0.01	0.01
25%	0.25	0.25
50%	0.33	0.33
75%	0.46	0.45
MAX	1.00	0.99

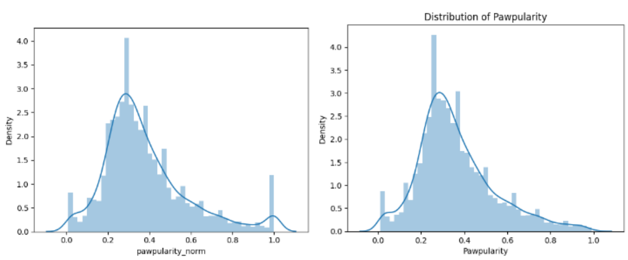

Figure 1 - Distribution plots.

C. *Model evaluation*

The current problem is considered a regression problem. This way, most of the models were trained using a Training/Validation approach, with a split of 20%, and some were trained using a 5-fold Cross Validation approach. Their performance was evaluated and compared utilizing the Root Mean Squared Error (RMSE) and the Coefficient of Determination (R^2) metrics.

D. *Metadata Analysis*

Firstly, we started by testing the predictive potential of the metadata features alone. This posed a challenge given that the number of features was small (12) and they belonged to a binary domain. To estimate which features could have the most predictive power, their distribution was analyzed through a bar chart (to check for unbalanced features) and a box plot (for assessing the distribution of the target variable in each group (0 or 1)). Afterwards, the target variable was binarized in order to perform an association rules analysis. Values of Pawpularity below the mean were considered as 0 and values above the mean were considered as 1. This method works with categorical or binary data to express relationships between variables. Given the low number of features, it was opted to analyze the frequency of each feature as an antecedent of the target variable on the 10 most confident rules and selecting the most frequent as the most relevant. Then, a comprehensive set of regression models were employed, including the Linear Regressor (LR), Decision Tree Regressor (DT), Support Vector Regressor (SVR), and Light Gradient Boosting Machine (LGB) Regressor, to analyze and predict the relationship between input features and the continuous output variable. All of these models were trained and selected for the dataset with and without outliers. This variety, inspired by similar problems in the literature and solutions of the Kaggle competition, allowed to assess if there was a more satisfactory model.

To explore the potentialities of a Neural Network approach in this issue, a MLP model was trained with a Grid Search Cross Validation to obtain the best parameters and later tested in the test set. The tested hyperparameters referred to the number of hidden layers (3 or 5), the activation function (ReLu or tanh), the penalty term for L2 Regularization and the learning rate, which could be constant or adaptative and have several initial values.

E. *Image Feature Extraction and Analysis*

In addition to utilizing metadata with binary features, which covered the apparent aspects we deem significant when assessing an image of a pet, we opted to analyze the images, by extracting features that encompassed patterns, textures, and colors – elements that influence our perception of images, even if not immediately obvious. Our objective was to validate our hypothesis, so we extracted 43 features and examined whether there existed a correlation with the Pawpularity score.

These features included Gabor, LPB and HOG, which were used in the literature to help describe the pose and facial texture of humans. (Kang Wang, 2015)(Timo Ahonen, 2004)(N. Dalal, 2005) Given the numerous quantity of HOG features, only the top 20 were used (when correlated with the Pawpularity score). Moreover, other references gave importance to image features, such as Haralick, Saturation, Entropy. Finally, we also extracted features related to Color, such as the mean of each RGB and its variance. The combination of these features may be helpful to quantify the details in a pet's fur, facial features, and assess the color vibrancy and image complexity, which can play a role in determining the aesthetic appeal of pet images. Secondly, we created a DataFrame "features.csv" with the normalized Pawpularity to facilitate model training and improve stability.

In total, five models were computed: Linear Regression (LR), Decision Tree Regression (DT), Support Vector Regression (SVR), Ridge Regression (RR) – λ = 0.01, 0.1, 1, 10, 100, 1000 – and Random Forest Regression (RF) – with 20, 30, 50 trees. For LR and RR there was an attempt to select the most relevant features in predicting the target variable, so Recursive Feature

Elimination (RFE) was applied (Brownlee, 2020), with the options of 8, 18, 28, 35, 43 features, randomly chosen.

To test a Deep Learning model in this type of data, a simple CNN model, composed by one 2D Convolutional Layer with a same convolution and a kernel of 3 pixels, followed by a ReLu activation function and a fully connected layer that computes the output target variable was performed. As input, the images were resized into size (3, 224, 224) and the model was trained for 100 epochs with a batch size of 32 and a learning rate of 0.001. Although this work aimed to produce other comparable Deep Learning approaches to this problem, the lack of computational resources at the time prevented from performing a thorough fine-tuning process, thus the results of this model must be taken into account as a predictor of the potential Deep Learning results with this problem.

F. *Combined Approach*

To combine both categories of features, the same types of models were trained with a combination of metadata and features extracted from the images. The same models as in section E were employed with the new training set.

III. Results

A. *Metadata*

From the distribution of the binary features, it was possible to conclude that all were severely unbalanced, with a large amount of images presenting the same value of each binary feature. Moreover, when analyzing the boxplots, it was possible to assess that the mean and standard deviation of the "Pawpularity score" did not differ depending on the value of any of the features. While there were slight deviations, no significant difference was visualized on both datasets. This also counter indicates the literature affirmation that pictures with blur, eye contact, and the proximity of the animal to the camera have major influence on the animal's popularity, at least in this case.

From the ranked list of features provided by the association rules, the LR model was trained with an increasing number of features and the best performing model on validation used 4 features in the full set and only 1 feature in the set without outliers. The DT and SVR model were trained with cross validation and the Light GBM (LGB) model was trained in a train validation setting. The test results for all these models are presented on Table 2.

Table 2 - Results of the Metadata models on the testing data.

		LR	DT	SVR	LGB
with outliers	RMSE	0.210	0.21021	0.21196	0.213294
	R^2	-0.0003	0.00037	-0.01638	-0.003114
without outliers	RMSE	0.2110	0.21111	0.21251	0.214286
	R^2	-0.007	-0.00825	-0.02167	-0.012465

Lastly, the best MLP model yielded the following results: RMSE = 0.2101 and R^2= 0.0016 for the model with a tanh activation, alpha = 0.01, 5 layers of 100 neurons each and a constant learning rate of 0.001.

Although similar, all the results obtained using this simpler type of data were unsatisfactory. The elevated values of RMSE, paired with the mostly negative or very low values of R^2 show that the model can barely predict better than guessing the mean of the dataset. This might show that these features are not enough to predict the value of the target.

B. *Image's Features*

The initial step after extracting and selecting the 43 features was to compute the correlation matrix, to check if the Pawpularity score had a significant association to any of these features. The results showed the highest correlation coefficient of 0.06, suggesting a notably weak relation between the analyzed features and the score, meaning there is limited evidence to support a strong linear relationship. Following this task, the five models were tested – the best results for each are shown on *Table 3* (for RR, α=100, and for RF, trees=30 and trees=20, respectively).

Table 3 - Results of the models for the testing data.

		LR	DT	SVR	RR	RF
with outliers	RMSE	0.2095	0.2104	0.2104	0.2094	0.2092
	R^2	0.0072	-0.0014	-0.0013	0.0079	0.0101
without outliers	RMSE	0.2056	0.2114	0.2115	0.2102	0.2135
	R^2	0.0029	-0.0106	-0.0121	0.0005	-0.0052

As observed on *Table 3* when comparing the overall results between the performance of the machine learning methods with and without the outliers, the values are analogous, although slightly better if executing the models with the outliers (both for the RMSE and R^2). In fact, the RMSE also revealed high values and the R^2 negative or low values, implying a weak prediction. It is also important to mention that the outcome of the training and testing sets were similar.

On the other hand, the efficacy of different regression algorithms varies based on the nature of the problem. LR without outliers and RF with outliers have exhibited finer results in the prediction of the Pawpularity score.

Regarding the feature selection, for the LR the best combination involved the usage of 35 features and as for the RR the employment of 18 and 28 features, with and without outliers, respectively. Among all the features selected, the ones that carried the most significance were Haralick and HOG features, so they had the highest impact in the analysis.

Lastly, in this topic, the CNN trained for 100 epochs, reaching a minimum in the validation loss at the last epoch. The result of the test RMSE from this model was 0.104233.

C. *Combined Approach*

When using both types of features, the same model regressors as before were employed. In the end, the models with the lowest validation loss were tested in the test set and the results obtained are presented in Table **4**.

Table 4 - Results of the models trained with both types of features for the testing data.

		LR	**DT**	**SVR**	**RR**	**RF**
with outliers	**RMSE**	0.21283	0.21325	0.21460	0.2135	0.21318
	R^2	0.00123	-0.0028	-0.0154	0.0057	-0.0021
without outliers	**RMSE**	0.21415	0.21545	0.21568	0.21363	0.21459
	R^2	-0.0112	-0.0235	-0.0027	-0.0063	-0.0154

IV. DISCUSSION

A. *Outliers*

The removal of the outliers for the model training was an insignificant effort as most results for the test data were either similar or worse. This might indicate that the hypothesis that the data followed a simple distribution close to a normal distribution is not correct. As a matter of fact, the characterization as a bimodal distribution could be more suitable and models that have into account such characteristics should be tested.

B. *Metadata*

With the analysis of the feature correlation with the target variable and the results of the model's regressions, it was confirmed that, when considering the problem of this particular website, the presence of certain features in an image is not distinctive enough to achieve a high Pawpularity. Given the unbalanced nature of the features, this might be due to the overall crowded environment of the website with many similar photos with similar elements.

A. *Image's Features and Analysis*

The correlation matrix has shown that the features extracted might not capture the full level of complexity involving the perception of an image with appeal, as cuteness of pets and image traits are often subjective and influenced by intricate visual cues that may not be adequately represented in a linear correlation analysis. To improve this aspect, the segmentation of the animals could be useful to capture features, but the dataset provided had very heterogeneous images, which would make it very challenging, as seen on ***Figure 2***.

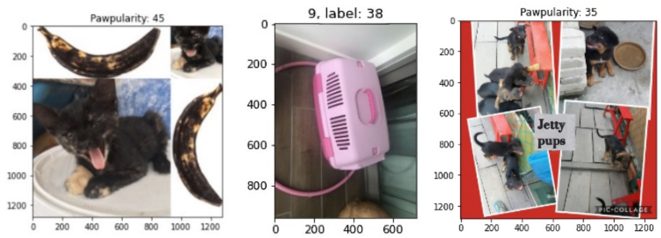

Figure 2 - Examples of photos in dataset.

Additionally, the values obtained for the training and testing set have suggest that the models are neither underfitting nor overfitting, meaning they have a consistent performance when faced with unseen data, although not very successful.

Furthermore, if the outliers are minimal, LR tends to produce precise results due to its sensitivity to the linear relationships within the data. Conversely, RF's robustness in handling complex and noisy data, allows it to capture non-linear patterns and outliers more effectively.

Now, further discussing the CNN performance, it is possible to note that, with this approach, the test loss was significantly lower than in other models. Furthermore, when compared to the known results of the original competition, it significantly outperformed the best models, that achieved a RMSE of 16.82256. This can be a very positive sign of the power of the deep learning approaches in this problem. However, it is still dubious that such a simple network could be better than several other more complex deep learning approaches on the Kaggle website. Therefore, this result is questionable and might be due to the unknown test set that the competitors faced when submitting their work, which might introduce more uncertainty on the test than with the data set that was used in this project. Another explanation could be an unknow occurrence of data leakage from the test set to the training set during the training process. Nevertheless, with the descending curve of training and validation loss presented in Figure 3 when training the model, it is possible to recognize the potential of discovering and integrating other more complex deep learning approaches for this dataset.

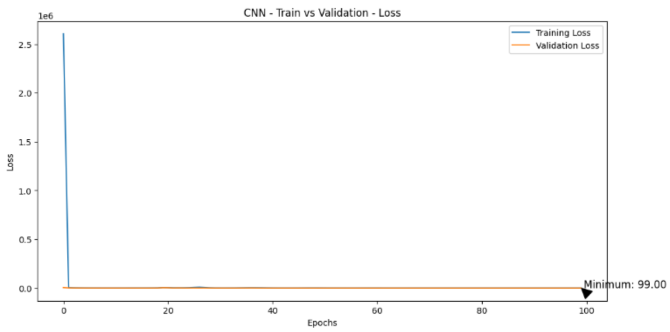

Figure 3 - Train vs Validation Loss of the CNN

B. *Overall models discussion*

The overall agreement is that all the machine learning models in all approaches delivered a very limited and unsatisfactory output, which suggests that the data used in these methods does not contain enough or representative information about the desired aspect of the pictures. Although following common and extensive procedures for computer vision feature extraction in images, the results suggest that the key features for determining what makes an animal appealing to the human eye cannot be translated by commonly used features. This way, and given that the opportunity for developing image based Deep Neural Networks hasn't presented itself in a more complex way in this work, due to lack of computational power limitations, and given that the prospects from the models that were in fact completed suggested an optimistic paradigm, the development of more robust, fine tunned and complex models using Neural Networks on image data is the main proposition for a better solution to this problem.

V. CONCLUSION

The analysis of the models concludes that the traditional Machine Learning approaches presented themselves short of the desired performance, translating the complex nature of what really makes an animal cute and appealing to the human eye.
Moreover, the diverse natures of the data on the dataset, shown by images with very different characteristics of animals performing in a wide range of "Pawpularity", add a complexity layer to this problem that could be further developed in Deep Learning or Ensemble Learning Approaches.

VI. REFERENCES

Brownlee, J. (2020). *Recursive Feature Elimination (RFE) for Feature Selection in Python.* Data Preparation.

Hiroshi Nittono, M. F. (2012). *The power of Kawaii: viewing cute images promotes a careful behavior and narrows attentional focus.* PLoS One.

How Photos Are Important To Pet Adoption: A Study. (2018). Obtido de Heartspeak_black_and_Red: https://heartsspeak.org/how-photos-are-important-to-pet-adoption-a-study/

Kang Wang, T. V. (2015). *Sense Beyond Expressions: Cuteness.* arXiv.

Melanie L Glocker 1, D. D. (2009). *Baby schema modulates the brain reward system in nulliparous women.* Proc Natl Acad Sci U S A.

N. Dalal, B. T. (2005). *Histograms of oriented gradients for human detection.* USA: IEEE.

PetFinder.my. (s.d.). Obtido de https://petfinder.my/

PetFinder.my - Pawpularity Contest. (23 de Sep de 2021). Obtido de kaggle: https://www.kaggle.com/competitions/petfinder-pawpularity-score

Pires, P. (Agosto 2023). Há quase 80 mil animais a viver em canis que ninguém quer. Solução pode ter de passar por esterilização obrigatória. *CNN Portugal*.

Timo Ahonen, A. H. (2004). *Face Recognition with Local Binary Patterns.* Finland: DBLP.

Witte, R. L. (2014). *Speed of Dog Adoption: Impact of Online Photo Traits.* Journal of Applied Animal Welfare Science.