Kaggle Pawpularity Contest: Predicting shelter pet popularity from profile images

Alex Teboul, Shea Dettling, Hima Spandana
Group 2

Abstract

This paper explores the use of machine learning models toward the task of predicting shelter pet popularity from human-generated metadata and raw pet profile images for Kaggle's PetFinder.my Pawpularity Contest. Previous work has demonstrated the effectiveness of Convolutional Neural Networks (CNNs) for image classification tasks. We show that supplying our convolutional neural networks with metadata information about the photo through concatenation methods can boost a CNN's predictive power. Additionally, we demonstrate the use of the pre-trained Xception model for boosting the predictive power of our models by adding a breed classification to training images and metadata. Methods covered include exploratory data analysis, the use of classic machine learning techniques for building models with the metadata, basic CNNs for image classification, mixed Artificial Neural Networks combining CNNs and metadata, and use of the pre-trained Xception model for pet species and breed detection. We discuss the shortcomings of the competition and why the Pawpularity score generated by the competition hosts cannot be accurately predicted from either the metadata or the raw pet profile images. We also share a tutorial series created for beginner Kaggle users to work through each step of the project.

Keywords: Machine Learning, Image Classification, Convolutional Neural Networks, Xception, Kaggle

Introduction

I. Background

Millions of stray animals suffer on the streets or are euthanized in shelters every day around the world. Our task was helping PetFinder.my, a Malaysian animal shelter, predict their trademark 'Pawpularity' score, which is an aggregate feature representing a pet's profile engagement from website visitors. This score, our sole dependent variable in question, ranges from 1-100, is designed to represent how often people interacted with an animal's profile. This metric serves as a proxy for how cute and adoptable the loving pet is. You might expect pets with attractive photos to generate more interest and be adopted faster, the ultimate goal of any shelter. In providing the shelter with a predicted Pawpularity score, we can help suggest when an animal's appeal is appropriately captured before making the profile picture live and ensure that the animal be united with a loving family as quickly as possible. But what makes a good picture?

PetFinder.my suggest that you take clear, sharp photos (not dim or blurry ones), to use a professional camera, use natural lighting, do not use a cluttered backdrop, have the animal doing in focus and doing an action, have the face centered and eyes in focus, among other useful tips. Many of these suggestions are what the human-annotated metadata was structured to track. The features available were all binary representations of the following 12 explanatory variables at our disposal: subject focus, eyes in focus, face prominence, nearness of the animal, action shot, accessories included, group photo, collage style, and human presence. The result was a tabular matrix of these 12 features to accompany the 9912 photos provided for training.

Submissions in the Kaggle competition are scored on the root mean squared error. RMSE is defined as:

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(\hat{y}_i - y_i)^2}{n}}$$

where $\hat{y}_i$ is the predicted Pawpularity value and y_i is the original value for each instance i. This metric provides a good indication of how close our model predictions for Pawpularity match the true Pawpularity labels on the test data.

II. Literature Review

Prior research has discussed the feasibility of regression problems using only binary variables, in a similar fashion to our metadata [1]. Geoffrey Tso and Kelvin Yau compared results between stepwise regression analysis, decision tree and neural networks in predicting electricity consumption for the Hong Kong municipal authorities. Their findings showed that all three types of models produced similar feature importances: number of household occupants, square footage of the household, and air conditioner ownership. In the summer models, the decision tree marginally performed the best, and in the winter model, the neural network performed best, showing the competitiveness of the model types. Our metadata models leverage similar techniques, though all produce similar RMSE metrics - highlighting the importance of the dataset and target variable for predictive power. Tso and Yau, however, help us distinguish that the premise of regression on binary features was not an ill-fated attempt but that the statistical properties of our metadata were what inhibited learning.

Andre Pacheco and Renato Krohling demonstrate how incorporating patient demographics when predicting skin cancer through images of skin lesions can improve accuracy [2]. They also proposed a novel attention based mechanism called the Metadata Processing Block (MetaBlock). It consists of a Neural Network layer that acts in a similar way to memory gates in Long-Term Short Memory (LSTM) models by selecting which information from feature maps of a CNN layer are retained. Results from two datasets showed that classification performance improved in 6 out of 10 scenarios. We will later demonstrate the usefulness of combining metadata into our CNN architecture but will not be implementing the MetaBlock layers.

Data augmentation is well established in computer vision research to improve the effectiveness of image classification models [3]. Specifically, simple techniques, such as cropping, rotating, zooming, and flipping input images have been shown to improve machine learning model classification accuracy on datasets such as Imagenet [3][4]. Both our Pawpularity contest dataset and Imagenet data sets contain pet images, so it stands to reason that data augmentation could improve classifier accuracy in our project as well. S. Wong et. al demonstrated that Convolutional Neural Networks in particular benefit the most from additional training samples provided by data augmentation, in that increasing the number of samples improves both the training error and the test error [6]. J. Wang and L. Perez at Stanford found that traditional data augmentation techniques are very effective at improving classifier accuracy, even when compared to more complex techniques like using CycleGAN to generate additional training images [4].

With the release of the Imagenet dataset roughly a decade ago, a variety of highly accurate image classification model architectures have emerged [5]. Convolutional neural networks (CNNs) and their derivatives have traditionally dominated this space. Beginning with LeNet-style models which use stacks of convolutions for feature extraction and max-pooling operations for spatial sub-sampling, CNN models have been increasing in complexity and accuracy [6]. Further improvements with AlexNet in 2012, where convolution operations were being repeated multiple times between max-pooling operations, allowed networks to learn richer features at every spatial scale [5]. One of the top models for Imagenet is the Inception class of model architectures. An improved Inception architecture called Xception, for “Extreme Inception”, was proposed in 2017. It offers similar accuracy, with greater efficiency by focusing entirely on depthwise separable convolution layers [7]. It achieves 79% Top-1 accuracy and 94.5% top-5 accuracy. This is the model architecture we use in our project for Breed identification.

Other recent advances in computer vision have seen the use of transformers used for image classification instead of convolutional neural networks [8][9]. The Swin Transformer architecture proposed in 2021 has shown particular promise as a general purpose backbone for computer vision applications [9]. It achieves 81.3% Top-1 accuracy and 95.6% top-5 accuracy on the Imagenet dataset. This is the model currently used by the top Kaggle leaderboard notebook for the Pawpularity contest. Other promising vision transformer models include Z. Jiang et. al's 2021 work on token labeling, the model can achieve 84.4% Top-1 accuracy on ImageNet [10]. These techniques are worth trying as future methods for improving over what we present in this paper. Ultimately these are model architectures that could improve our ability to predict Pawpularity. It is already well established that similar models can predict dog breed [11], so it stands to reason that these models could extract relevant features from the pet images provided in this Kaggle competition.

Methods

I. Exploratory Data Analysis

Within the metadata was the true Pawpularity score for the corresponding photograph. The data provided had no missing values or inputs of the wrong type, thus no cleaning was needed for the metadata - recall it was hand-built for this project. The underlying distribution was quite skewed with fat tails. This (see Figure 1) normal quantile-quantile plot highlights how asymmetrical our distribution was with both tails being unusually saturated.

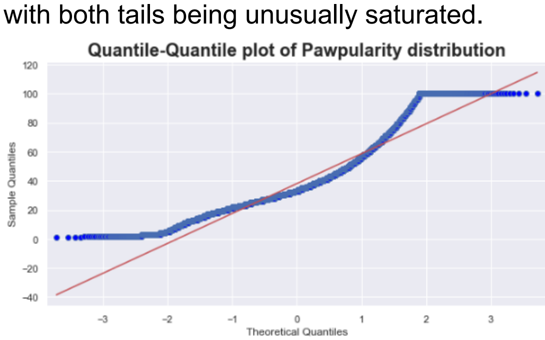


Figure 1

The mean score was roughly 38, however, the majority of scores spanned the 20s and 30s with a sizable portion of the entire distribution being situated right at the right bound of 100 (see Figure 2).

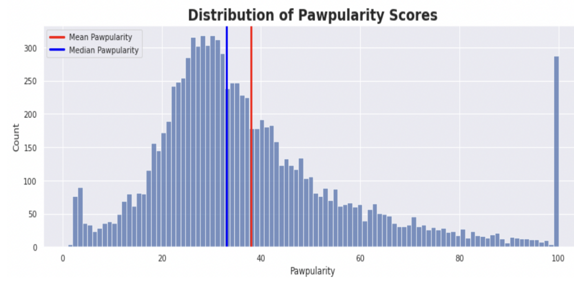


Figure 2

During the model building process, we discovered that predicting the upper end of scores proves quite challenging and getting our prediction to follow such an atypical distribution proved difficult at first for our first few iterations of CNNs.

Another issue with the metadata was that it essentially amounted to a sparse matrix. There were only three columns that had mean scores above 0.5 (see Appendix 1 for a full statistical summary), indicating more 1s than 0s were presented. However, for the remaining features we rarely see any instances of 1s. For example, only 0.9% of all 9912 photos had instances of the animal in action, and 2.7% for subject focus, both recommendations that PetFinder.my stated as useful techniques. Quickly, it will become clear what a challenge is posed to discriminate Pawpularity when the typical metadata observations consist of the same three features regularly; two of those three features had variable inflation factors (VIFs) above 10 (see Appendix 3), which indicates that they are highly collinear with other features in the dataset. A full correlation matrix can be found in Appendix 2; it highlights the largest correlation between any explanatory and dependent variables at under 2%. The lack of collinearity between explanatory variables also points to the fact that many of the same photos largely ignored the guidelines on good photo taking, hindering nonlinear discimrination by Pawpularity score. This was a continuing pattern as we moved to predict Pawpularity score solely based on the metadata values and found that prediction was severely stunted by the sparse nature of the data.

II. Metadata Model Building

In many models, the effort to minimize any loss function was largely done by predicting the mean to get the lowest error variance. Take our Stochastic Gradient Descent Regressor, for example, (results in Figure 3) that was severely handicapped by the lack of diverse scores and honed in on the mean, rarely predicting anything beyond 40.

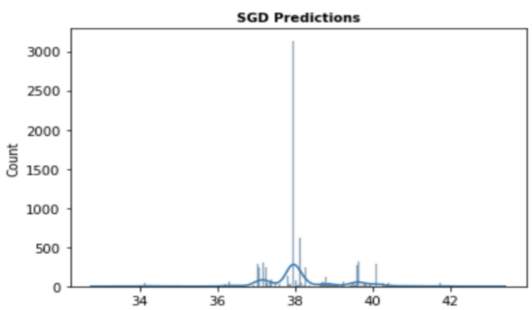


Figure 3

This was one of many attempts at non-linear regression that ultimately led to the same result. KNN regressors, Decision-Tree regressors, and Random Forest Regressors had much the same issue. In Appendix 4, a table shows the relative similarity of scores for a variety of ensemble methods. Even attempts to transform our matrix in preparation of these techniques did not yield the expected results. Non-negative matrix factorization (NMF), Kernel tricks, and Principal Component Analysis (PCA) were all conducted in the hopes to alter the existing dimensions of the matrix in a way that would flesh out a new axis along which ML techniques could learn but all attempts led to the same mean predicting patterns. Perhaps it was wishful to hope that 12 binary features would accurately describe a task as complicated as predicting how an emotion-driven human would engage with the photographs of adorable animals.

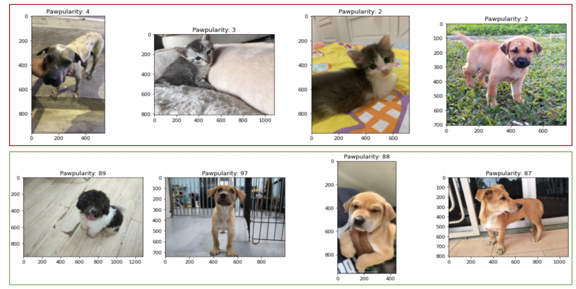


Figure 4

We were already aware that there seemed to be some inconsistencies in the Pawpularity scores when judged by their photos alone. You must keep in mind that the Pawpularity score is an aggregated score representing far more underlying statistics about the digital engagement with the animal's entire profile. There could be many contributing factors that are inaccessible for the sake of this research that might have shed light on why some animals would be receiving scores below 10 when they are subjectively no less cute than their counterparts in the shelter. Take Figure 4 as an example of the disparity in scores between animals that are captured remarkably similar fashion. This can also been seen in a different manner through a series of duplicate photos that were detected using a 90% threshold on cosine similarity. Figure 5 illustrates how duplicate profiles had notably different scores, begging the question of what is really driving engagement that is either not being captured or masked underneath the aggregation leading up to a Pawpularity score.

Figure 5

To summarize, we created a tutorial for metadata model building in which we explore Decision Tree Classification, Decision Tree Regression, Ordinary Least Squares

Regression, Ridge Regression, Bernoulli Naive Bayes Classification, Random Forest Regression, and Histogram-based Gradient Boosting Regression (LightGBM). The results of which can be seen below in Figure 6. All of these models fail to accurately learn from the metadata and simply end up predicting the mean Pawpularity score for nearly every new image. The sparsity of the metadata and the lack of relationship between the Pawpularity score and these 12 human generated metadata features is to blame for the poor performance.

Model	RMSE
4.1 Decision Tree Regression	20.857
4.2 Decision Tree Classification	22.900
5.1 Ordinary Least Square Regression	20.827
5.2 Ridge Regression	20.827
6.1 Bernoulli Naive Bayes Classification	23.468
7.1 Random Forest Regression	20.838
7.2 Histogram-based Gradient Boosting Regression	20.924

Figure 6: Metadata Models

III. Convolutional Neural Networks

In preparation for training our first CNN, analysis was conducted on the dimensions of the training photos, which were not standardized in any manner. It was found that the most common photo maintained a 1.5 aspect ratio (see Figure 7 for dimension histograms), indicating the vertical nature of most photos. It was determined this was due to the apparent number of photographs taken on digital phones with the camera behind held upright.

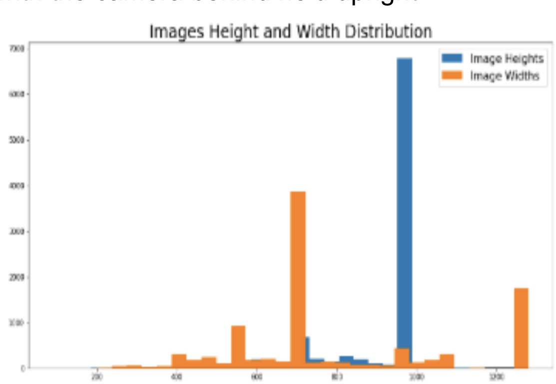


Figure 7

Thus, the photos were reduced to 48x72 to allow for the least amount of tear possible when compressing the images. The color values were customarily normalized down to a range of [0,1] and the appropriate channel dimensions were added to support the 3-dimensional nature of any encoded photo. A test-train split of 30% was used for the prediction visualizations coming up, in addition to a 20% further validation split to monitor generalization throughout training. Our first model consisted of two convolution layers with a stride of one, 3 by 3 kernels, Rectified Linear Units (ReLu) as the activation function followed by two fully-connected layers with and ReLu activation again. Batch normalization, a process of standardizing the data, either direct inputs of the activations of a layer, by fixing the mean and variance, which, in practice, helps stabilize results and improve speed. Dropout was also applied at 25% to the fully connected layers and the CNN layers after max pooling.

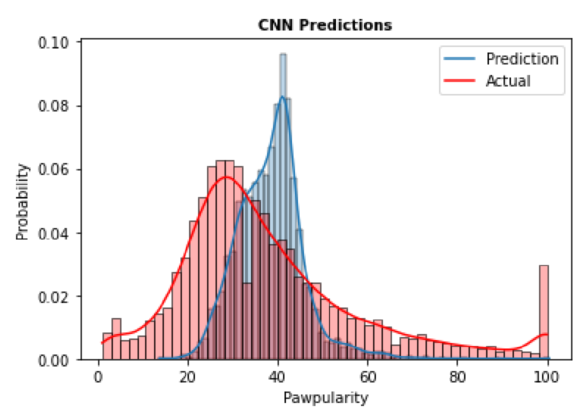


Figure 8

The distribution of predictions can be seen in Figure 7. This will be a recurring visualization to check the levels of fitting that our model produces. The red histogram represents the distribution of all the scores for the full 9912 data points; the blue histogram represents the produced distribution of predictions, which should nearly match the 'true' distribution.

The important thing to note is the degree to which we are able to reproduce the same oddly-shaped distribution of scores instead of just checking MSE since we had seen earlier how greedy models will go for the low-hanging fruit of predicting the mean to lower error variance. These are much more intelligible predictions than with any of the metadata models, however, the shape of true distribution still eludes us. The validation MSE in the last epoch of training was 21.27 (Model 1). The thought at this stage was that the huge spike of scores at 100 was dragging the prediction distribution to the right so some quick attempts were made to restructure the true distribution.

Smoothing out any score of 100 randomly across the range of 90-100 did not improve scores, nor did entirely removing any score observation of 100 from the training set help alleviate the skew of the predictions. Thus, any further training manipulation was discarded as the lack of results did not justify our greedy approach. Sticking with the original training set, a grid search was then implemented to see how far we could push accuracy.

After tuning across the number of dense nodes, activation function, loss metric, optimizer, convolution layer depth, and convolution map depth we were still not seeing much improvement to MSE; the last epoch has a validation MSE of 23.07 and the underlying prediction distribution was still not satisfactory (see Figure 9 and Model 2).

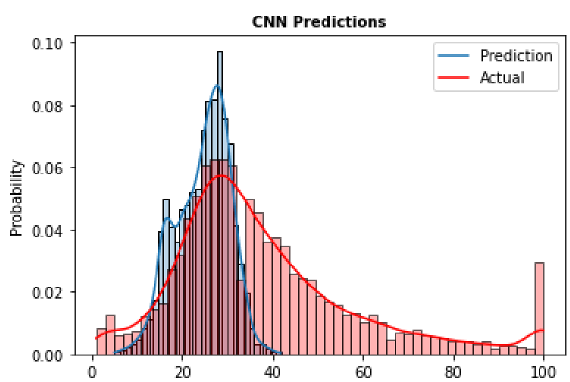


Figure 9

It was at this point, having felt we exhausted the usefulness of typical CNN parameters and standard architectures that we decided to integrate the metadata and images into one coherent model. Two architectures were proposed (see Figure 10). They both consisted of much the same for the CNN portion of the network; one combined the predictions of the CNN and the prediction from a two-layer ANN on the metadata, which fed into another two-layer ANN as a two-feature input (left side Figure 10); the second network also contained the same CNN structure but placed the prediction into the metadata matrix as an additional column, which was then fed as a 13-feature input to a slightly deeper ANN to account for the lack of a third ANN.

Figure 10

Test scores were marginally worse at 25.7 MSE (Model 3). However, the interesting aspect of this model was it's added complexity allowed it to overfit to a degree, which actually finally allowed our predictions to mimic the true distribution of Pawpularity scores (see Figure 11). So while the test MSE has not improved, one could argue that the model has finally conformed to the training data.

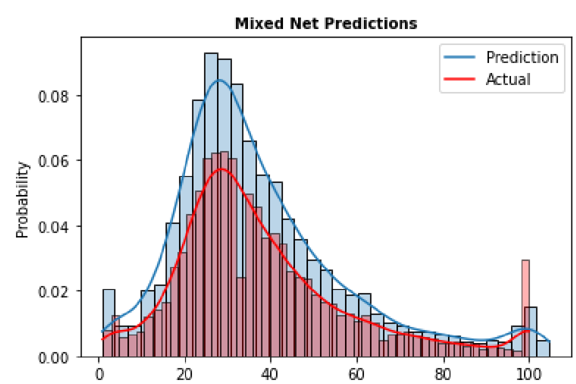


Figure 11

The second model (on the right) from Figure 9 had similar scores with a validation MSE of 25.13 (Model 4). Adding kernel regularizers and data augmentation in the form of random flipping and a 20% random contrast adjustment, improved our validation MSE slightly up to 23.79 (Model 5). The minimal change possibly points to the fact that our model had a fair amount of generalization naturally baked into it from the varied nature of the training photos discussed earlier; people were, quite literally, submitting photos of their animals in all shapes and sizes. Despite the first CNN (Model 1) having a lower validation MSE, this regularized mixed net (Model 5) will be considered the best model thus far since it is better equipped to accurately describe the distribution of scores, which is far from normally distributed.

IV. Xception Breed Classification Models

Given the limited success achieved with the metadata models and basic convolutional neural networks, it was clear that whatever signal existed to predict pawpularity from these pet profile images was very limited. In order to determine whether different species/breeds showed different mean pawpularity scores, we turned to a pre-trained model: Xception. Xception has been trained on the Imagenet dataset to predict over 1000 classes. Without performing transfer learning, or fine tuning, we implemented Xception to predict pet profile image classes, calling the result Breed. This method led to a number of misclassifications. Misclassifications were defined as image classifications falling outside of dog or cat species/breeds. Misclassified image breeds were replaced by a 'misclassified' breed label, which ended up having a mean score equal to the mean Pawpularity score of the whole dataset at 38.

We applied this Xception breed classifier to the entire training dataset and then took the mean Pawpularity score for each breed (see Figure 12 for a demo).

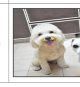
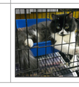
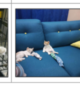
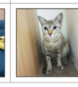
Predicted Breed	Maltese dog	Golden Retriever	Persian cat	Studio couch	Egyptian cat	Boston Bull
Average Pawpularity	86.8	52.9	52.4	39.3	30.9	27.0
Example						

Figure 12

As seen in Figure 11, we observe the promising result that different breeds classified by the Xception model have on average different mean Pawpularity scores. Based on this knowledge, we mapped the mean pawpularity score for the predicted breed classification to all training images and performed the same on previously unseen images in the test data. If the Xception model predicted a previously unseen class, a pawpularity score of 38 was assumed. We call this added data feature 'BreedBoost' versus the classification itself which we refer to as Breed.

Next, we built a random forest model using the prior metadata plus Xception predicted Breed. Dummy variables were created for each breed classification. This resulted in improved RMSE over just the metadata alone. The new RMSE was 19.85 versus 20.84 without the Xception Breeds. To further improve the models, we ran this again with the average Pawpularity scores by breed, 'BreedBoost', mapped in as well. This further improved RMSE to 18.74. So the native Xception model was able to bring a simple Random Forest ensemble to 19.85 with Breed classifications and 18.74 RMSE with the addition of the BreedBoost feature. This method resulted in our top RMSE when submitting to the Kaggle competition test data.

Conclusions

The task of reducing validation RMSE proved more challenging than anticipated but we would argue that the process of learning continued in a straight path, nonetheless. We had much higher hopes for the metadata being able to offer insights into which photo taking guidelines would boost profile engagement but the lack of positive examples proved too great a challenge. We went from metadata models largely predicting the mean to our first CNNs ironically having the lowest test scores but rarely predicting scores above 40. As network complexity was increased and hyperparameters tuned, we were able to dramatically lower training scores in a way that showed the extra degrees of freedom were allowing the distribution to be accurately described. However, the trouble was test scores were not improved; in many cases, better training scores led to worse test scores. This was not seen as detrimental during our model evaluation, since we had really reached a prediction distribution that nearly mirrored the unlikely distribution of Pawpularity scores.

The final piece of the puzzle was incorporating pre-trained models that detected the species of the animals. There were a handful of misclassifications outside of any animal species. Given more time, we would have structured the Xception predictions to take the highest probability dog or cat species instead of the most likely label given all possibilities. It was clear and not unexpected that certain breeds were given preference on the site, something we suspect the shelter was already aware of.

Regardless of the different model architectures tried in this project, it is clear that the Pawpularity score has little signal to it. In other words it is difficult to use either the images themselves or metadata based on the images to predict the Pawpularity score. There is more that goes into pet selection on the website and click rates for profiles than just the features of the images.

Author Contributions

Hima Spandana was responsible for exploratory data analysis and metadata model building. Shea Dettling built many of the early CNN and constructed the networks that allowed for incorporating metadata and images. Alex Teboul made our tutorial series, improved and tuned our CNNs, pushing the networks further by applying pre-trained models that distinguished animal species. All authors performed each stage of the project process individually from EDA to Model building to gain experience, but came together to share results and define best practices.

Results

Image Models	Train RMSE	Test RMSE
1 : First CNN	7.43	21.27
2 : Grid Searched CNN	22.79	23.07
3 : Mixed CNN + Metadata Pred	1.99	25.57
4 : Mixed 13-input Metadata	1.80	25.13
5 : Regularized Mixed Net	3.09	23.79
6: Random Forest Metadata + Xception Breed	-	19.85
7: Random Forest Metadata + BreedBoost	-	18.74

Appendix

Image 1 :

	Subject Focus	Eyes	Face	Near	Action	Accessory	Group	Collage	Human	Occlusion	Info	Blur	Pawpularity
count	9912.000000	9912.000000	9912.000000	9912.000000	9912.000000	9912.000000	9912.000000	9912.000000	9912.000000	9912.000000	9912.000000	9912.000000	9912.000000
mean	0.027643	0.772599	0.903955	0.861582	0.009988	0.067797	0.129338	0.049637	0.166263	0.172014	0.061239	0.070420	38.039044
std	0.163957	0.419175	0.294668	0.345356	0.099444	0.251409	0.335591	0.217204	0.372335	0.377411	0.239780	0.255866	20.591990
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	1.000000
25%	0.000000	1.000000	1.000000	1.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	25.000000
50%	0.000000	1.000000	1.000000	1.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	33.000000
75%	0.000000	1.000000	1.000000	1.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	46.000000
max	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	100.000000

Image 2 :

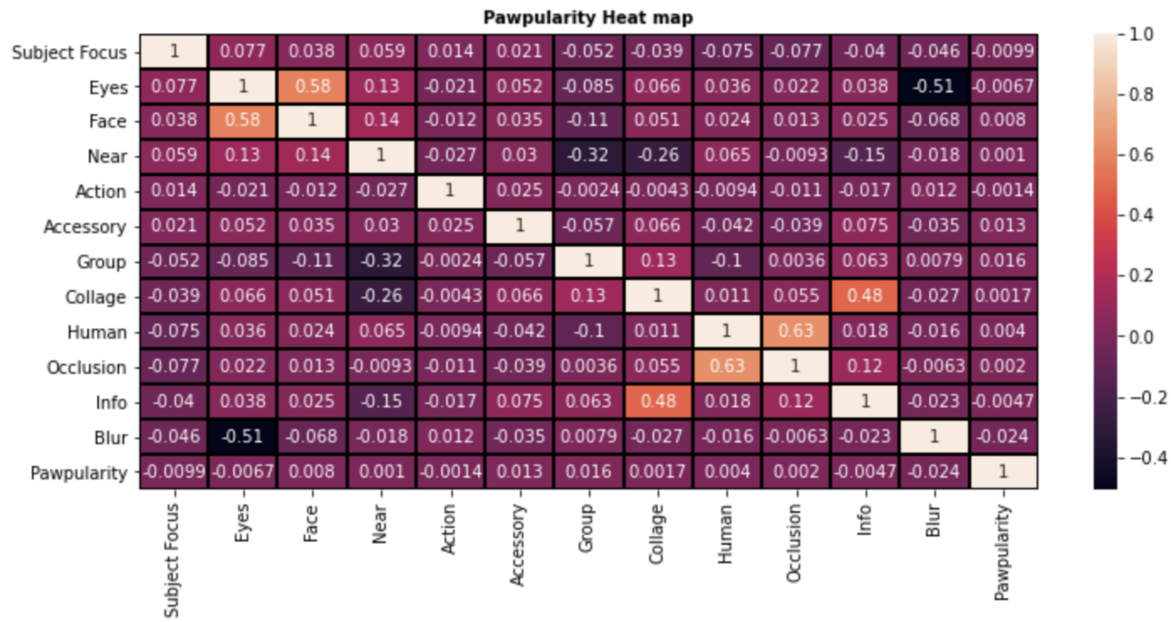


Image 3 :

	feature	VIF
2	Face	13.715668
1	Eyes	10.118170
3	Near	5.762924
9	Occlusion	2.073562
8	Human	2.064939
11	Blur	1.595109
7	Collage	1.452023
10	Info	1.412621
6	Group	1.163850
5	Accessory	1.090942
0	Subject Focus	1.048292
4	Action	1.010174

Image 4:

	mse	reg_score	test_score	train_score
AdaBoost	23.738432	0.051251	-0.321623	-0.247558
Bagging	20.901994	0.087167	-0.024658	0.035581
Bagging & AdaBoost	22.026695	0.062954	-0.137895	-0.066537
DecisionTree	21.118528	0.085956	-0.045998	0.037946
GradientBoost	20.663094	0.085149	-0.001370	0.016442
LinearRegression	20.655592	0.083535	-0.000643	0.003090
RandomForest	20.900269	0.087167	-0.024489	0.035562
XGB	21.053791	0.087974	-0.039595	0.037527

References:

[1] Geoffrey K.F. Tso, Kelvin K.W. Yau, Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks, Energy, Volume 32, Issue 9, 2007, Pages 1761-1768, ISSN 0360-5442,
https://doi.org/10.1016/j.energy.2006.11.010.

[2] A. G. C. Pacheco and R. A. Krohling, "An Attention-Based Mechanism to Combine Images and Metadata in Deep Learning Models Applied to Skin Cancer Classification," in IEEE Journal of Biomedical and Health Informatics, vol. 25, no. 9, pp. 3554-3563, Sept. 2021, doi: 10.1109/JBHI.2021.3062002.

[3] Wong SC, Gatt A, Stamatescu V, McDonnell MD. Understanding data augmentation for classification: when to warp?. In2016 international conference on digital image computing: techniques and applications (DICTA) 2016 Nov 30 (pp. 1-6). IEEE.

[4] Perez, L., & Wang, J. (2017). The effectiveness of data augmentation in image classification using deep learning. arXiv preprint arXiv:1712.04621.

[5] Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems. 2012;25:1097-105.

[6] Y. LeCun, L. Jackel, L. Bottou, C. Cortes, J. S. Denker, H. Drucker, I. Guyon, U. Muller, E. Sackinger, P. Simard, et al. Learning algorithms for classification: A comparison on handwritten digit recognition. Neural networks: the statistical mechanics perspective, 261:276, 1995.

[7] Chollet F. Xception: Deep learning with depthwise separable convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition 2017 (pp. 1251-1258).

[8] Han K, Wang Y, Chen H, Chen X, Guo J, Liu Z, Tang Y, Xiao A, Xu C, Xu Y, Yang Z. A survey on visual transformer. arXiv preprint arXiv:2012.12556. 2020 Dec 23.

[9] Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., ... & Guo, B. (2021). Swin transformer: Hierarchical vision transformer using shifted windows. arXiv preprint arXiv:2103.14030.

[10] Jiang Z, Hou Q, Yuan L, Daquan Z, Shi Y, Jin X, Wang A, Feng J. All tokens matter: Token labeling for training better vision transformers. In Thirty-Fifth Conference on Neural Information Processing Systems 2021 May 21.

[11] Wang C, Wang J, Du Q, Yang X. Dog Breed Classification Based on Deep Learning. In 2020 13th International Symposium on Computational Intelligence and Design (ISCID) 2020 Dec 12 (pp. 209-212). IEEE.