Cat Breed Classifier

# Description of the problem

**Brief description of the problem and its context (industry, environment, etc.)**

Cats were domesticated around 10 000 years ago. They had played an important role in ancient civilizations, such as Egypt. Over the years hundreds of different cat breeds were developed, all over the world, using a method known as selective breeding. In the past they were treated as a means of pest control, while nowadays most cats are kept as pets. This project will focus on seventeen breeds, which are most popular in the United States.

**Who can benefit from this model, and how?**

This project aims to benefit cat lovers, who don’t know breed description by heart. Some breeds are closely related to each other and therefore also very similar. This model may help to differentiate breeds such as, for example, Thai and Tonkinese cat.

**Why does the problem seem interesting to you?**

I had several cats since I was seven. I also really liked the task with frog classifier we did during the classes. It differentiated between three different species of frogs. I was wondering how hard it is to train a machine learning model, which will be able to handle close to twenty categories.

# Data

**Sources of data, assessment of its reliability**

In order to gather training data, I utilized built-in method of the fast-ai library - search\_images\_ddg.  
As a parameter I passed class name, which is the name of breed, common keyword – ‘cat’ and number of requested pictures (in my case 200). The method searches pictures using the engine DuckDuckGo and afterwards downloads the images to directories of the specific classes. The data is not 100% reliable, sometimes a photo is labeled incorrectly, or it can’t be opened. It is however, more than enough, for the specified task.

**Brief descriptive analysis of the data**

The training data is divided into directories, which serve as a label for the photos. In each directory there is close to 200 pictures of cats, representing a specific breed. The photos are very diverse, they are taken from different angles and in different lighting. Sometimes only part of a cat is shown, there may be few cats in one picture or some other objects. Images depict adult cats, as well as kittens.

**Rationale: how can this data help solve the problem?**

The data provided gives the model a detailed outlook of the selected breeds. Basing on it the model should be able to differentiate between breeds, when faced with new pictures.

# Solution

**Brief description of the chosen model with justification**

For this project I picked a Convolutional Neural Network (CNN) model, generated using PyTorch and FastAI libraries. I used pre-trained network called resnet34. It was trained previously on a very large (100 000 + images) dataset named ImageNet. Resnet34 is considered to be a state-of-the-art model for image classification. The fact that it is pre-trained allowed me to use a methodology called transfer learning, which greatly improves the quality of a model trained on relatively small amount of data.

**Stages of project implementation**

Firstly, the training data is fetched using the previously mentioned – search\_images\_ddg method. Then it is split randomly into training and test data sets in 80/20 proportions. Then the data is augmented by resizing, cropping and rotating using the methods provided by the FastAI DataBlock API. Then a batch of data is shown in order to verify, whether the previous steps were executed correctly. Then learning-rate finder is called, in order to optimize the base\_lr hyperparameter. After, the model is trained. By the means of trial and error I found out that the best results are provided with 6 epochs and base\_lr hyperparameter equal to 2e-3. Finally, a confusion matrix is displayed, along with most confused categories and the model is exported into a pickle.

**Measures of evaluation (quality assessment) of the model**

I used two main quality assessment metrics:

* Macro-Average F1 Score (simply an average F1 scores of different classes)
* Error-Rate (which is simply 1 – Accuracy)

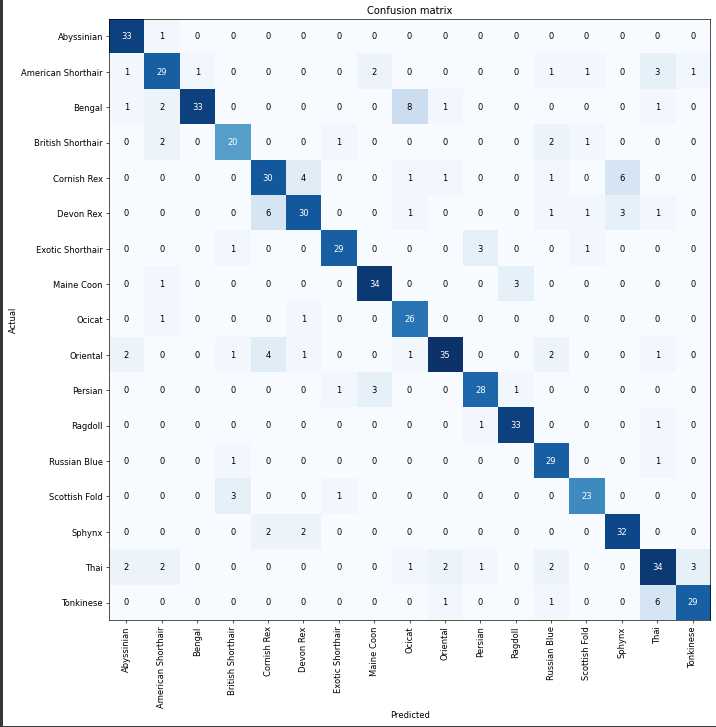
# Discussion of Results and Model Evaluation

**Modeling results**

The trained model was put into a StreamLit web app in which user can upload a picture of a cat and a breed will be determined with a provided probability. The app was then manually tested with pictures taken offline.

**Model evaluation**

Below you can find the confusion matrix, error\_rate and macro F1-score. I believe that the F1-Score of 0.8161 is solid, regarding the number of classes and the quality of the data. Most of the incorrect results are between breeds, which are closely related and therefore similar to each other



# Obraz zawierający tekst Opis wygenerowany automatycznie

# Reflections

**What was successful?**

The app based on the model works well. I tested it on pictures of cats belonging to my family and friends and it was correct in all but one case, in which it confused Thai cat with Ragdoll breed. You can find couple of test pictures attached.

**What were the problems? How did we solve them?**

The main challenge was finding an optimal number of epochs and ways of training the model. I tried few approaches such as using different combinations of fit\_one\_cycle method calls and utilizing discriminative learning rates (as described in fastbook tutorial), but in the end the basic approach, using the fine\_tune method worked out best.

**How can this be used in the future?**

The model can be further expanded, by adding more breeds and perhaps by trying the more advanced resnet50 pre-trained model. It would require more resources to train, but it may be able to provide more precise results.

# References

While I was preparing the project, I used materials provided by the teacher during the classes. I also used online sources listed below:

<https://github.com/fastai/fastbook>

<https://docs.fast.ai/>

<https://models.roboflow.com/classification/resnet34>

<https://peltarion.com/knowledge-center/documentation/evaluation-view/classification-loss-metrics/macro-f1-score>

<https://parade.com/1275429/stephanieosmanski/most-popular-cat-breeds/>