



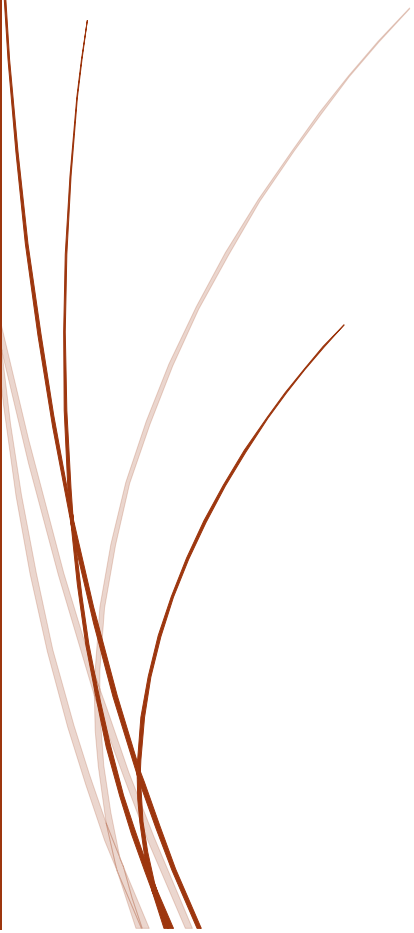
# Dog Breed Classifier

Capstone Project Proposal

Machine Learning Engineer Nanodegree

Ahmed Karawia

July ~ 2021



## Contents

Domain Background .....	2
Problem Statement .....	2
Datasets and Inputs .....	2
Solution Statement.....	2
Benchmark Model .....	2
Evaluation Metrics.....	3
Project Design .....	3

## Domain Background

Identifying dogs according to breeds is a crucial issue for dogs' buyers, sellers, and even for veterinarian doctors. Each dog breed has certain physical and behavioural characteristics<sup>i</sup>. Furthermore, knowing the breed will provide information such as future size and required care. These all at the end are reflected on the price and value of your dog.

Having a handy solution such as an application which can identify the breed will improve and facilitate this task. This project shows a machine learning model which can predict the breed when provided a dog image and even it can predict the most resemble breed when the provided image is for human utilizing Convolutional Neural Networks (CNNs) technique.

## Problem Statement

This project provides a solution to predict dogs' breed when provided with a dog image and it can predict the most resemble breed when provided with a human image.

It involves the utilization of Convolutional Neural Networks (CNNs) technique to build image classifier to achieve this goal.

## Datasets and Inputs

In this project the following datasets provided by Udacity will be used:

- Dog dataset contains 8351 images.<sup>ii</sup>
- Human dataset contains 5750 images.<sup>iii</sup>

All inputs including training, testing and validation are of image type to full fill the goal.

Dog dataset has 8351 images (6,680 Images for training the model, 836 Images for testing, 835 Images for validation). Number of images provided for each breed is not the same which leads to imbalance and there is variation in sizes and background among the images.

Human dataset has 13233 images, and all have the same size 250x250 but have different backgrounds.

## Solution Statement

It involves the following steps:

1. Detect human using OpenCV's with Haar feature based cascade classifiers.
2. Detect dog using pretrained VGG16 model.
3. Predict the breed: classified images are passed to CNN to return the best breed match.

## Benchmark Model

The benchmark model will be according to the Kaggle leaderboard for dog breed<sup>iv</sup> less than 0.01 which is in the top 100 of the competition. Another benchmark would be 90% accuracy.

## Evaluation Metrics

The evaluation metrics will be as illustrated by the Kaggle dog breed identification competition as multi-class log loss between the target and the probability which is between 0 and 1. The goal is to minimize the loss (ideally to zero).

## Project Design

The project is designed according to steps provided by the standard template as follows:

- **Step 0:** Import Datasets and required libraries.
- **Step 1:** Detect Humans using OpenCV's implementation of Haar feature-based cascade classifiers.
- **Step 2:** Detect Dogs using pre-trained model (Pre-trained VGG-16 Model)
- **Step 3:** Create a CNN to Classify Dog Breeds (from Scratch).
- **Step 4:** Create a CNN to Classify Dog Breeds (using Transfer Learning).
- **Step 5:** Writing the Algorithm.
- **Step 6:** Testing the Algorithm.

---

<sup>i</sup> Editorial, P. M. D. (2021, February 15). *Importance of Purebred Research*. PetMD.  
[https://www.petmd.com/dog/general-health/evr\\_dg\\_importance\\_of\\_purebred\\_research](https://www.petmd.com/dog/general-health/evr_dg_importance_of_purebred_research).

<sup>ii</sup> <https://s3-us-west-1.amazonaws.com/udacity-aind/dog-project/dogImages.zip>

<sup>iii</sup> <https://s3-us-west-1.amazonaws.com/udacity-aind/dog-project/lfw.zip>

<sup>iv</sup> *Dog Breed Identification*. Kaggle. (n.d.). <https://www.kaggle.com/c/dog-breed-identification/leaderboard>