# Project Proposal: Hummingbird Classification

# Amanda Kimball – akimball002@regis.edu

# Regis University

# May/June 2020

## Forward:

For my Practicum project, I am choosing to evaluate a non-business or work related problem. I would like this effort to be publicly displayed and critiqued so that I can better my skills as a data scientist. The ultimate goal of this work will be to create a supervised classification model that can be used to predict the gender and species of a hummingbirds from a video, photograph and/or text description of the bird in question.

## Data Collection:

My initial task will be to collection of already classified hummingbird photos from the web for the initial models used in the classification. I will augment them with photos from my own collection as needed to obtain a reasonable dataset. It is my hope to accumulate 100+ images for each of the species/gender segments in the dataset. The data will be collected and stored in a file structure as displayed below:

Figure 1: Directory structure for hummingbird photographs dataset.

Reportedly, there are over 330 species of hummingbirds with 115 genera [1]. However, I plan to stick to the ones native to Colorado which is reportedly 5 [2]: Black Chinned - Archilochus alexandri, Broad-tailed - *Selasphorus platycercus*, Costa - *Calypte costae*, Calliope - *Stellula calliope*, and Rufous - *Selasphorus rufus* for a modest 1000+ photo dataset.

The key to successful image classification will be to initially create an image cleanup tool. Each image used in the dataset will need to be adjusted show primarily the hummingbird. Ideally a animal detection model could be built to pick out the bird in question and then crop the jpg photo to show mostly the bird. Once the photo is cleaned it can be added to a dataset (training, test, validation). It would be ideal to have this same model to process future photos that need predictive classification (more test samples). The 190-birds dataset on Kaggle has 2 hummingbird data subsets that can be used for this initial pre-processing cleanup algorithm training [3].

## Data Analysis:

An initial binary classification model to break the birds into male and female gender groups. This is an interesting and important task as the various female bird species are generally less colorful than their male counterparts. I will split the data into training, testing, and validation test sets. Create sequential model for determination of the gender. I’ll use augmentation, pre-trained models and various other neural network classification tools to build an optimal model, fit the entire training dataset to the binary classification model with validation dataset used for optimization of the accuracy. Finally I will test the model against the test dataset and see how well the model predicts the correct gender of the birds.

A second – categorical model architecture will be used to classify the 5 individual species. The male dataset will be fit separately from the female dataset, but it is my hope that the architecture will work for both. The same general method will be used for the second model architecture except it will a 5 classes rather than being binary.

## Project Risk Assessment:

Finding adequate pictures and cleaning the data to uniform model inputs will be the highest risk portion of this project. There is no clean hummingbird dataset that I’m aware readily available for modeling.

Project Timeline:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Task Week | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| Initial Project Proposal |  |  |  |  |  |  |  |  |
| Collect/classify photos from the web and my camera |  |  |  |  |  |  |  |  |
| Build data cleaning / hummingbird image processing tool |  |  |  |  |  |  |  |  |
| Create binary male/female classification model |  |  |  |  |  |  |  |  |
| Create multi-categorical species classification model |  |  |  |  |  |  |  |  |
| Draft Final paper / GitHub repository / Kaggle uploads |  |  |  |  |  |  |  |  |

## References:

[1] <https://www.hummingbirdcentral.com/hummingbird-species.htm>

[2] <https://www.beautyofbirds.com/hummingbirdscolorado.html>

[3] <https://www.kaggle.com/gpiosenka/100-bird-species/version/23>

[4]

Further ideas to consider:

Image classifier app: <https://towardsdatascience.com/detecting-cute-animals-with-machine-learning-d39a511bd144>