Advanced Data Science Project Proposal

Team Members: Susan Nunez, Claire Robinson, Madison Yonash

Project Idea: Cat or Dog Detection Model and Implementation through a Shiny App

https://www.kaggle.com/datasets/karakaggle/kaggle-cat-vs-dog-dataset

Data description (codebook, data dictionary): Include a table or list with the name of the attributes being used, their type, and their description.

This data set includes 24,000 images that are a subset of three million annotated images compiled by Petfinder.com and Microsoft. The classes are balanced as there are 12,000 images for each class. This dataset does not come presplit into training and testing sets, but they have been separated based on whether they contain a dog or a cat. As an added challenge for the model, many of the images have additional noise or obstructions in front of or near the animal. Lastly, the images in the dataset are not uniform. In other words, they vary in size, and some are in landscape while others are portrait orientation

Below are some examples of the images with an example of both obstructed and unobstructed views:

	Cat	**Dog**
Unobstructed		
Obstructed		

Goal: a revised (if needed) version of the main project goal and a clear statement on how the dataset you chose is relevant/useful to achieve such goal.

For our final project, we plan to create an image classifier that uses images of cats and dogs. We will use the above dataset. We plan to use neural networks to accomplish this task, implemented in Python. There are many examples of creating a neural network to perform this task, so we will continue to research different examples and use them as references when creating our model. This model will then be used in a Shiny application, where a user can upload a photo and the app will output the result of the model, if the uploaded image is classified as a dog or cat.

References: related references (blogs, research papers, studies, etc.) that you have reviewed so far for the specific topic of interest you chose.

Brownlee, J. (2021, December 7). *How to classify photos of dogs and cats (with 97% accuracy)*. MachineLearningMastery.com. https://machinelearningmastery.com/how-to-develop-a-convolutional-neural-network-to-classify-photos-of-dogs-and-cats/

An example using a similar cat and dog dataset of photos to build neural networks for classification using Python. It will be beneficial to reference in terms of how to start the modeling process and look for different ways we can make the project our own.

Callens, A. (n.d.). *Shiny_Classifier: Shiny application to manually classify images by pushing buttons.* GitHub. https://github.com/AurelienCallens/Shiny_Classifier

Gives an example of how to create a Shiny app for a classification task, will be useful in determining the UI we are creating and how to implement the server versus UI.

Deploying a shiny app with a tensorflow model. TensorFlow for R. (n.d.). https://tensorflow.rstudio.com/guides/deploy/shiny.html

This is an example of implementing Tensorflow into a Shiny App. It can be useful because it uses a Keras API as the model. It also provides another example of creating a Shiny App and other advanced models that can be applied into the app.

Lendio, L. (2023, September 6). *Cats vs dogs image classification model*. Kaggle. https://www.kaggle.com/code/orensa/cats-vs-dogs-image-classification-model

An image classification model that shows the different steps when creating and implementing the model. This was an entry in the Kaggle competition that is where we are getting our data set.

Padhiar, K. (2020, May 1). *Cat vs dog dataset*. Kaggle. https://www.kaggle.com/datasets/karakaggle/kaggle-cat-vs-dog-dataset

This is the dataset that contains the cat and dog photos that we will be testing on our Shiny App.

Questions: (if any) please list any questions or concerns you may have at this point regarding the final project and the progress made so far (e.g., difficulty cleaning the data, the need to understand how to encode certain attributes, doubts about the modeling technique you plan to explore - if any -, coding errors/issues)

We have yet to get started on the project but are excited to undertake it. No members of the team have worked on image classification or neural networks before, so this is a new endeavor for all of us. Additionally, figuring out the preprocessing steps that are necessary both for training models on images and uploading images for classification will be fun to learn. Our team also has limited Shiny experience, so we are looking forward to adding another skill to our data science repertoire.