WGU EDU

Task 2

Cat/Dog Image Classification System

Chase Christensen

Student ID # 001528190

10/10/2022

Contents

[A.1. Letter of Transmittal 3](#_Toc116311940)

[A.2 Project Proposal 5](#_Toc116311941)

[Problem Summary: 5](#_Toc116311942)

[Application Benefits: 5](#_Toc116311943)

[Data Outline: 5](#_Toc116311944)

[Data Description: 6](#_Toc116311945)

[Objective and Hypothesis: 6](#_Toc116311946)

[Project Methodology: 6](#_Toc116311947)

[Funding Requirements: 6](#_Toc116311948)

[Stakeholder Impact 6](#_Toc116311949)

[Ethical Considerations: 7](#_Toc116311950)

[Developer’s Expertise: 7](#_Toc116311951)

[B: Executive Summary 8](#_Toc116311952)

[Problem Statement, Customer Summary and Current System Analysis: 8](#_Toc116311953)

[Data: 8](#_Toc116311954)

[Project Methodology: 8](#_Toc116311955)

[Project Outcomes: 8](#_Toc116311956)

[Implementation Plan: 9](#_Toc116311957)

[Evaluation Plan: 9](#_Toc116311958)

[Resources and Costs: 10](#_Toc116311959)

[Timeline and Milestone: 11](#_Toc116311960)

[D: Post Implementation Report 12](#_Toc116311961)

[Purpose: 12](#_Toc116311962)

[Dataset: 12](#_Toc116311963)

[Data Product Code/Data Visualizations/Breakdown of Accuracy: 14](#_Toc116311964)

[Testing and Robustness: 17](#_Toc116311965)

[Summation: 17](#_Toc116311966)

[Appendices: 18](#_Toc116311967)

[Included files: 18](#_Toc116311968)

[Installation Guide: 18](#_Toc116311969)

[Prerequisites: 18](#_Toc116311970)

[Final Word of Learning Experience 20](#_Toc116311971)

# A.1. Letter of Transmittal

Joseph Fitzer,

Operations Manager

Raining Cats and Dogs

555 Highway Rd.

Murray UT, 84123

Dear Mr. Fitzer,

I think that we can both agree that what you do for cats and dogs, our beloved furry friends, is a gift to our society. But I can understand that the sheer number of animals that pass through your care can be daunting to manage. Between taking them in, housing then, feeding them, managing their health and temperament, and finally finding them a forever home is a hefty process that takes you and your employees’ full attention. As such, we have a potential solution to help quicken the process you face in managing these many as-of-yet unadopted pets. This can be done by an image classification system, which will let your database know, based on an uploaded image, what type of animal it is. This will let your employees upload images all at once, and rather than spend several minutes marking the type of animal, the database will do it for them, cutting time spent on data management.

To demonstrate this in action, we’ve collected over 20,000 images of cats and dogs. There’s a sizable variation in breed, environment, and positioning of the animals. We’ve gone ahead and created a model to show the efficacy of the system and its ability to classify images and sort them appropriately. There will be a graph included later that shows the growth of this program over time, showing that it does indeed become more effective as time continues until accuracy is over 99%. Of course, this can be scaled up to include pets in the rodent, reptile, avian, and other animal families.

This project will cost roughly $30,000 for development to deliver a working product. With our expertise, including my own history in software development across multiple languages, it is easy to promise a working image classification system that would perform to your expectations within the span of 50 business days.

If you would like to speak more on this, I would be happy to take some time to sit down with you and talk about any details and answer any questions you have. Feel free to reach out to me directly.

Sincerely,

Chase Christensen

Chief Engineer

# A.2 Project Proposal

## Problem Summary:

Raining Cats and Dogs is a humane society dedicated to the caring and adoption of common household pets. Their focus is on cats and dogs. Currently they are intaking many cats and dogs who are unwanted, wild, or pets of owners who can no longer care for them. Due to their focus being adoption, they take pictures to post on their social media and put into their database for ease of identification. However, right now that means an employee must take the picture, upload them to their database, and then sort these images manually. This is a step that takes a sizeable amount of time, time that could be better spent elsewhere. An image classification system could help cut out this step entirely via automation.

## Application Benefits:

The image classification system we’re proposing would save employee time. It would also provide an easy way to upload images to a website or social media location by allocating animal picture to their corresponding categories. It would also help remove any incorrect images, such as a picture lacking an animal entirely. Finally, this is easily scalable, and any additional animal type, such as rabbits or birds, would be easily included in this classification system.

## Data Outline:

The software will be developed in the Python programming language using premade libraries such as Tensorflow, Keras, numpy, all within the JupyterLab environment. The program is partitioned into smaller portions. From importing libraries, to importing the images, resizing the images, to running the train and test programs. Finally, a small visual aid is provided to show the growth of the program over time.

## Data Description:

The dataset used is a publicly available collection of pictures including a 50/50 split of cats and dogs. These images were obtained from Kaggle.com and all images were checked to ensure that they were the correct image type, preferably a .jpg file. Any image that does not meet this image requirement is not included in the application. Upon meeting this criterion, the images are then resized to a 150x150 ratio, recolored to red, green, blue color schemes, and set to a float32 format.

## Objective and Hypothesis:

The main objective of this image classification program is to reduce the amount of time that employees spend cataloguing new pets brought into their care. Raining Cats and Dogs already spends the time to sort the picture of their currently cared for pets. The hypothesis is that if this system is implemented then employees will spend less time managing the database and more time focusing on the animals, and more time spent getting pets adopted.

## Project Methodology:

We will use the Agile methodology called ‘Scrum’ to develop this system. And the platform used for this process will be JupyterLab, a system that allows for easy incremental product development. Since scrumming is segregated into sprints, this platform that will allow incremental design will be ideal.

## Funding Requirements:

Initial funding costs have been calculated at $30,000. This will cover development, deployment, and initial troubleshooting while employees are introduced to the system. However, this final number may change based on approved alterations in scope or timeline adjustments.

## Stakeholder Impact

The end goal of this process is to reduce employee overhead. Therefore, it will allow employees to spend more time caring for pets or working towards getting pets adopted. This will hopefully create a greater throughput from when animals are taken in to when they are adopted.

## Ethical Considerations:

Due to the primary target being animals, there are little to no security or ethical constraints as no sensitive data will be processed through this system.

## Developer’s Expertise:

Over the past two years, developers have completed projects using Java, Python C++, and SQL. These projects have included an inventory management system, an automated chatbot, a route calculator for delivery vehicles, and an appointment system. In particular, the automation and inventory management system will be very applicable in knowledge on how to best implement the system that is currently being proposed.

# B: Executive Summary

## Problem Statement, Customer Summary and Current System Analysis:

Raining Cats and Dogs has been known as Murray’s primary humane shelter for cats and dogs. At any given moment they have pets numbering in the hundreds with dozens being left in their care and another several dozen being adopted each day. As such their database is highly fluctuating with pets and their pictures. They have one employee whose entire job is simply to take pictures of the pets for their database and social media and then must input these new pets into that database and move adopted pets into a different ‘adopted’ database. Raining Cats and Dogs wishes this employee to spend less time on managing the database and more time on their social media and adoption efforts.

## Data:

The current dataset used for training the image classifier is a collection of cat and dog images publicly available on Kaggle.com. Links provided:

* <https://www.kaggle.com/datasets/shaunthesheep/microsoft-catsvsdogs-dataset>

Images will automatically be adjusted by the system. The user will not be required to do so themselves.

## Project Methodology:

Project will be developed using the Scrum Method within the Agile Methodology. A Scrum Master will be assigned to keep the project focused on the current sprint. The project is divided into several sprints, each taking roughly 40 business hours to complete. The Scrum Master will meet with the Dev Team each morning to discuss the current Sprint goals for the day as well as any concerns, complications, or approved changes in scope.

## Project Outcomes:

Product deliverables will include completed training datasets, the image classifier, and a system that will allow the image classifier to work with Raining Cats and Dogs current hardware and software.

## Implementation Plan:

At the outset the project will be using the Scrum framework. The main deliverables will be the image classification, the training image datasets, as well as a basic guide for users. Mr. Fitzer will be the primary contact for Raining Cats and Dogs, as well as the product owner. Tasks to complete a sprint will be assigned to the Dev Team at the beginning of each sprint in the morning meeting. Review of the previous sprint will be done at the closing of the related sprint, as well as a discussion on things that went well and parts that will need improvement for the next sprint. Upon completion, the software will be provided to the customer as Software as a Service.

## Evaluation Plan:

Due to the nature of JupyterLabs, evaluation and testing will be run incrementally as development continues. All work will be compared to acceptance criteria, as well as readability. Due to the agreement to provide troubleshooting and training, a high priority will be placed on controlling bugs and ease of use.

## Resources and Costs:

Current costs are listed at $30,000. The breakdown is as follows:

|  |  |  |
| --- | --- | --- |
| Type | Cost | Explanation |
| Hardware | $0 | Program will run sufficient on the customers current hardware (Windows 10) and no hardware changes are needed. |
| Software | $0 | All software (Python 3.9, Tensorflow, Keras, JupyterLab) are publicly available. No purchases necessary. |
| Labor | $30,000 | This system will take a total of 200 hours to complete as currently planned. This cost also includes troubleshooting and training for the first 3 months after install. |

## Timeline and Milestone:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Milestone | Start | End | Duration (hrs.) | Contributors | Description |
| Initialization | 10/17/2022 | 10/17/2022 | 8 | Product Owner, Dev Team, Scrum Master | Scrum Master Selected/Backlog developed |
| Sprint 1 | 10/18/2022 | 10/25/2022 | 40 | Scrum Master, Dev Team | Libraries imported. Dataset acquired, imported, and reviewed. |
| Sprint 2 | 10/26/2022 | 11/02/2022 | 40 | Scrum Master, Dev Team | Learning model developed |
| Sprint 3 | 11/03/2022 | 11/10/2022 | 40 | Scrum Master, Dev Team | Model training and Data display |
| Sprint 4 | 11/11/2022 | 11/18/2022 | 40 | Scrum Master, Dev Team | Model Test and User sections completed. |
| Final Product Delivered | 11/21/2022 | 11/21/2022 | 8 | Product Owner, Scrum Master, Dev Team | Product delivered to owner |
| Training and Troubleshooting | 11/21/2022 | 3/03/2023 | Up to 16 man-hours | Product Owner, Scrum Master | Remaining 16 hours used for training or troubleshooting. |
| Project Review | 11/22/2022 | 11/22/2022 | 8 | Scrum Master, Dev Team | Review with Dev Team. Input taken and project components saved. |

Total Time: 200 hours

# D: Post Implementation Report

## Purpose:

The purpose of this project was to create an image classifier that would be able to differentiate between cats and dogs. The project included training images and instructions will be given below if a user wished to test their own images. Visualizations of how the program performed were also provided for a breakdown of how the program can perform.

The idea of this program was to provide proof of concept that the program would serve Raining Cats and Dogs as a sorting system. A full build out would sort the images into folders that they can upload into batches they could then upload into their database. This would, in theory, help them cut down on employee time spent, automating at least one step of their process and save both time and money.

## Dataset:

The dataset was obtained from Kaggle.com and a link to the actual dataset was provided previous in this document. But for clarity, Figure 1 is a selection of 25 images that the program would see on running:

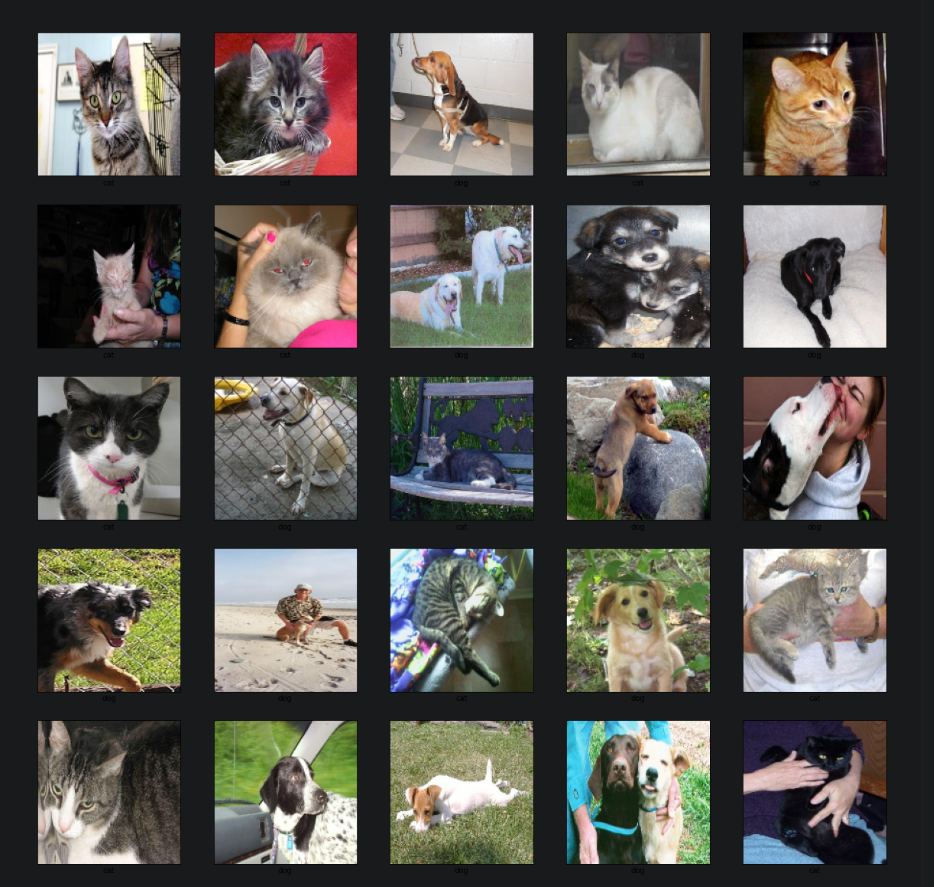


Figure 1

As one can see, the images might be just dogs or cats, multiple animals, or perhaps a human being included. All images were vetted beforehand for similar formats. If an image was erroneous, the program would not count it in the classifier via the usage of a simple for loop.

## Data Product Code/Data Visualizations/Breakdown of Accuracy:

The code itself was rather straightforward using libraries such as Tensorflow and Keras. First, the data was loaded as displayed in Figure 2:

Text

Description automatically generated

Figure 2

A total of 2,000 images were used for training. And to ensure equal numbers of training options a bar graph is used to display. The exact number of images is less important than an equal distribution of both cats and dogs. As we can see in Figure 3, the breakdown is very similar.

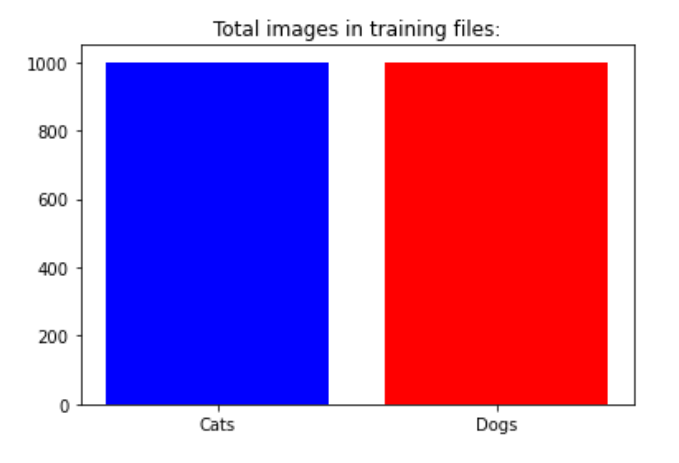


Figure 3

After it has loaded the images, it would then be sent through the training process. The training process only went through 10 epochs, though more could undoubtedly increase the accuracy before a true plateau is hit. Below is a visualization of the increase in accuracy from the first epoch, to the last with Figure 4 displaying epoch 1, and Figure 5 showing epoch 10.

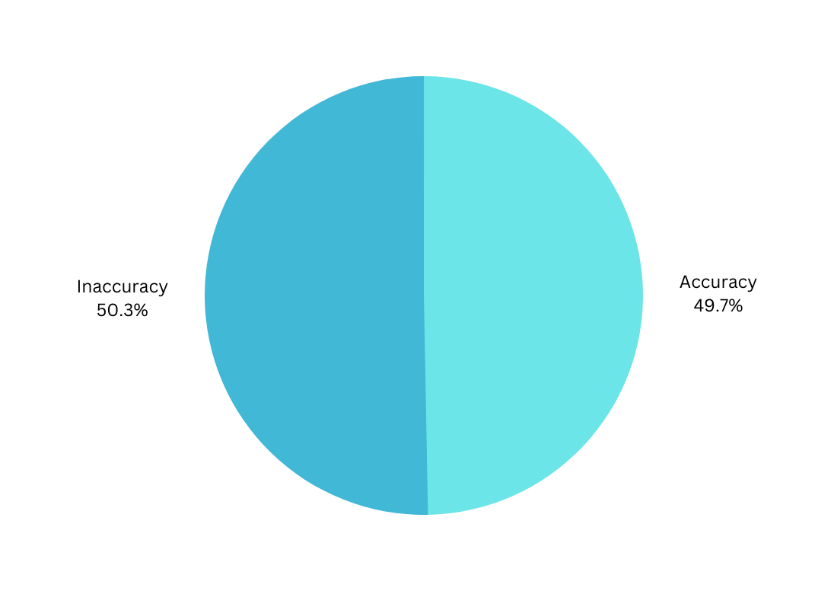


Figure 4: Accuracy of Epoch 1

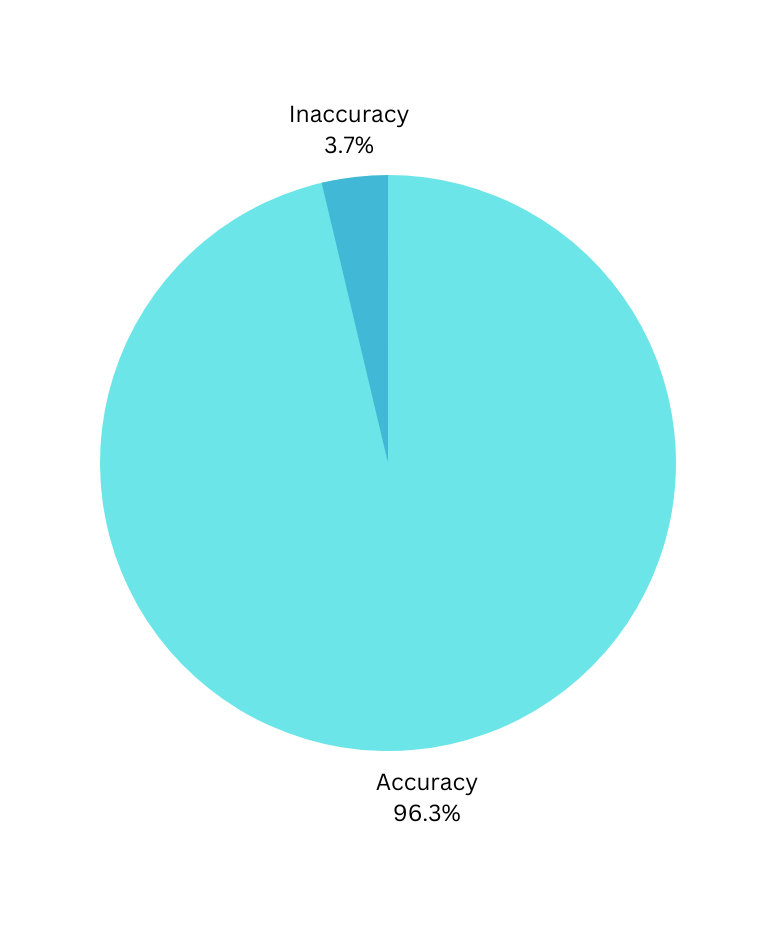


Figure 5: Accuracy of Epoch 10

Finally, to show to efficiency of the program itself, we can look at a breakdown of the increase in accuracy. However, this also shows the diminishing returns that are incurred when running multiple epochs. Overall, the cost per growth in accuracy increases drastically after Epoch 1. See Fig 6

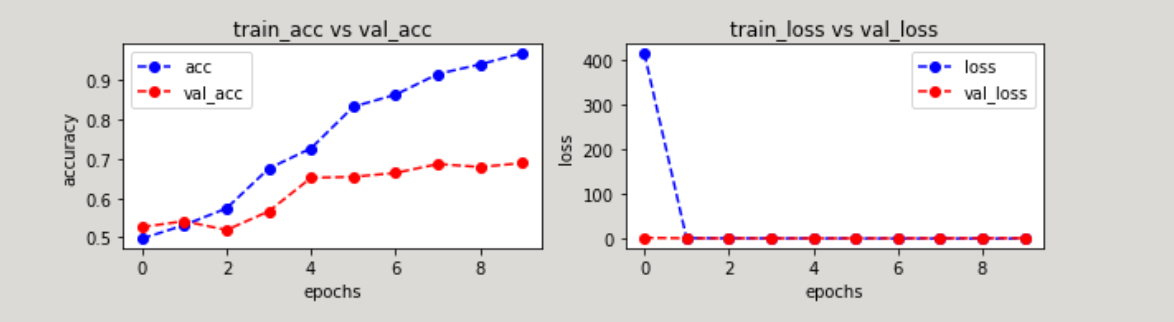


Figure 6

## Testing and Robustness:

All tests were performed via Scrum methodology. Each unit of Code was tested and proved fully functional by integration testing until all the parts could be brought together for the final build. No bugs have been found.

## Summation:

Overall, the program runs as intended. Though additional epochs of training are required before an ideal accuracy is reached. However, 96% accuracy is well within margins to save an employee a fantastic amount of time via automation brought about by the classification of images. However, it would be idea to build into the program the *sorting* of the images, and not just the classification. Eventually a full program could be used to automate the classification, sorting, and uploading of the images into Raining Cats and Dogs’ database.

# Appendices:

## Included files:

The following files are required for complete operation of the program.

CatsandDogs.ipynb is the JupyterLab Notebook file for the image classifier.

User\_Guide.pdf is the installation guide for the program. For convenience, that same information has been provided below.

microsoft-catsvsdogs-dataset: This includes the requisite images for testing. Within that file are two additional folders: Train and Test. Train is the focus of this program, and it contains 1,000 cat and 1,000 dog images. Additionally, there is a test folder with 100 images of cats and 100 images of dogs. This latter folder is not necessary unless one wishes to see how the program runs on untested images. This is already done at the very end of the program.

## Installation Guide:

### Prerequisites:

Python 3.9 and JupyterLab with supporting libraries: Keras, pandas, numpy, sklearn, cv2, Tensorflow, Matplotlib

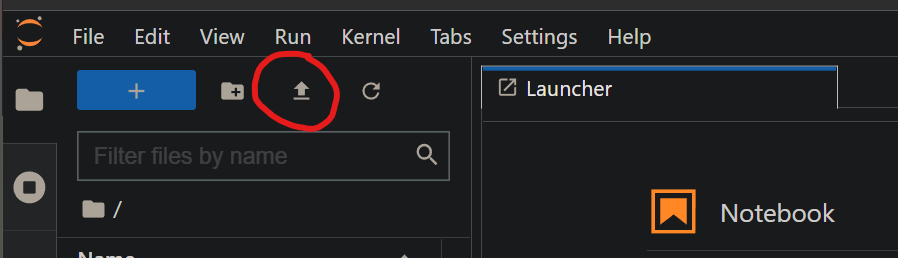
1. For easy installation install Anaconda from their website <https://www.anaconda.com/>.
2. Once Anaconda has been installed, navigate to the newly installed Anaconda Navigator
3. Be sure that Jupyter Notebook and JupyterLab is installed.
   1. When Anaconda Navigator is launched, check the environments included. Double check that the additional environments mentioned above are installed. (Keras, pandas, numpy, sklearn, cv2, Tensorflow, Matplotlib).
4. Launch JuptyerLab
5. Download and unzip the provided zip file.
6. Click the upload file found at the top left of the JupyterLab window that has now opened. (See Fig. 7)

Figure 7

1. Upload the two files located in the provided zip file named ‘CatsandDogs.ipynb’ and ‘microsoft-catsvsdogs-dataset’
2. Launch the CatsandDogs.ipynb from with JupyterLab by double-clicking on the uploaded file.
3. Press the ‘restart the kernel, then re-run the whole notebook’ button (see Fig 8). This may take some time for the program to run through to completion.

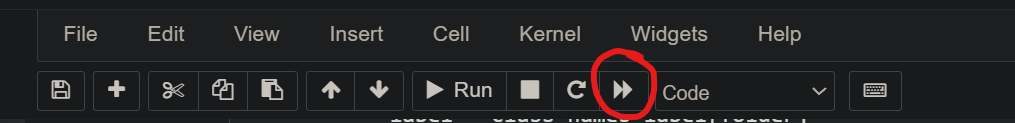


Figure 8

1. Throughout the running process you will see progress made such as progress bars, two data visualizations populated, and a final operation on a test file. This shows how the program runs on additional images it has received no training on.

## Final Word of Learning Experience

As I finish up my Computer Science degree through WGU, I feel like this project is fitting of the name capstone. What I mean by that is that this entire process through this college experience has been filled with learning how best *I* learn. And my ability to find answers, to learn and expand my current knowledge, has been refined to culminate in this final project. The Internet, course instructors, other training materials through third party education sources have all helped me understand how to best tackle this. And to be honest, I wasn’t sure how I would do it. It was overwhelming. Imposter syndrome reared its ugly head several times. Could I do this? Could I be successful? Ultimately, as I submit this final project for what I hope is the final time, I hope to find the answer to be a resounding yes.