**WGU C964**

**Computer Science Capstone Task 2**

**Barnyard Animal Image Classification via Convolutional Neural Network for Targeted Animal Feed Marketing**

**Jefferson Warnimont**

**Student ID #001482272**

**03-20-2023**

**Table of Contents**

Part A: Project Proposal for Business Executives 4

[Letter of Transmittal 4](#_Toc1044087116)

Project Proposal 6

Problem Summary 6

Application Benefits 6

Application Description 6

Data Description 6

Objectives 7

Methodology 7

Funding Requirements 8

Data Precautions 8

Developer’s Expertise 8

Part B: Project Proposal for Technical Staff 9

Problem Statement 9

Customer Summary 9

Existing System Analysis 9

Data 10

Project Methodology 10

Project Outcomes 11

Implementation Plan 11

Evaluation Plan 12

Resources and Costs 12

Timeline and Milestones 13

Part C: Application 15

Part D: Post-implementation Report 17

Project Purpose 17

Datasets 18

Data Product Code 19

Objective Verification 24

Effective Visualization and Reporting 24

Accuracy Analysis 25

Application Testing 25

Application Files 25

User Guide 27

[Summation of Learning Experience 28](#_Toc1702287338)

References………………………………………………………………………….……….....30

Part A

**Letter of** **Transmittal**

March 19, 2023

Jane Hunt, Owner, and CEO

ABC Feed Supply

12345 Feed Supply Rd.

Feedtown, Ohio

Dear Ms. Hunt,

I am writing you today after having discussed an exciting business prospect with your company’s website manager, Mike Jones. First off, congratulations on your company’s recent successes. In the course of our discussion, Mr. Jones informed me of the recent expansion of your business and the subsequent need to make improvements and upgrades to your website. We all know that a customer facing website is necessary for any successful business in today’s marketplace as consumers are more connected than ever to their devices and the technology they deliver and far less likely than in the past to do their shopping at brick-and-mortar locations. Thus, it is more important than ever before to have a web presence that is eye-catching, engaging, and informative in order to attract and maintain customer focus and promote brand recognition.

I believe a key feature of procuring and retaining customers is through interactive engagement. Through an innovative and feature rich platform we can maintain customer focus and attract new customers as well. An interesting idea that Mr. Jones and I discussed for an improvement in this area would be adding a feature that allows a customer to upload a photo of their pet or livestock animal and automatically receive a recommendation for one or more of your quality feed products. It is my understanding that your company produces feed formulas for eight types of barnyard animals including cats, dogs, chickens, cattle, sheep/goats, pigs, horses, and rabbits. I am confident that my development team can produce an artificial intelligence driven data product to perform the task of classifying the animal in a customer’s submitted photo to one of these eight classes with a high rate of accuracy and within a timeframe of six weeks from the acceptance of this proposal. Mr. Jones and his team could then integrate the solution into your website layout with their chosen web functionality features added.

I fully believe this site feature will turn heads and lead to the procurement and retention of customers. This in turn will lead to higher sales and profits. Customers will have the benefit of retrieving relevant information in a quick and efficient manner. This undertaking can also have the added benefit of encouraging more robust customer profiling which could be used to harvest data and judge market conditions for future expansions or targeted marketing campaigns. Once integrated into your website, this feature will be a godsend for your marketing department!

The baseline funding required to develop this data product will be a onetime development fee of $15,000. The data product itself shouldn’t require any maintenance and any website maintenance and operation costs would be included in Mr. Jones’ annual budget. I have received a Bachelor’s of Computer Science from a respected university, am versed in several programming languages and development platforms, have obtained multiple technical certifications related to development and project management, and have been a member of the professional workforce for over a decade. My team members have varying levels and modes of experience and are all industry professionals of the highest degree. Your project is in good hands with us!

Thank you for your consideration. I will be looking forward to your response and am happy to answer any questions or offer any clarifications as needed.

Sincerely,

Jefferson Warnimont

Senior Engineer

XYZ Development Inc.

**Project Proposal**

**Problem Summary:**

In today’s marketplace, it is a necessity for a company’s web presence to match or exceed its physical presence. Consumers are more connected to their devices and screens and less connected to their communities and local businesses. Online purchases are continually increasing and in-store, brick-and-mortar purchases are continually declining. Consumers have less time and attention-span than ever before and crave engagement, excitement, and convenience. ABC Feed Supply has recently expanded the operations of its physical business and its web presence needs an upgrade to keep up. The website manager has expressed a need for more engaging features that will grab and hold the attention of current and future customers.

**Solution and Benefits:**

XYZ Development Inc. is proposing the development and implementation of an artificial intelligence driven data product which, when viewing a submitted customer photo, will identify the type of animal pictured. This animal classification will be implemented in a website feature used to automatically recommend one or more appropriate products available from ABC Feed Supply. We believe the addition of this feature will lead to increased site traffic, which in turn will lead to higher levels of customer acquisition and retention and lead to higher sales volume. This feature will also create opportunities for more targeted marketing efforts both directly as its main functionality and indirectly in its potential for better data farming for use in analysis of markets and customer needs when it leads to customers adding more information to their profile.

**Application Description:**

The data product we will provide will be an artificial intelligence image classifier model trained on a dataset of images of the categories; Cat, Chicken, Cow, Dog, Goat/Sheep, Horse, Pig, and Rabbit. We will use a convolutional neural network (CNN) and utilize transfer learning and fine-tuning techniques on the popular CNN model MobileNet to attain a high level of accuracy when predicting among these eight animal classes with our own model. This model will be developed in the Python language on open-sourced software, and we will utilize image data freely available on Kaggle.com as well as images scraped from Google image searches to build a dataset specific to our needs for the training, testing, and validation of our model.

**Data Description:**

We will be utilizing publicly available images of cats, chickens, cows, dogs, goats/sheep, horses, pigs, and rabbits for training, validating, and testing the artificial intelligence model and these images will be sourced from Kaggle.com as well as Google image searches. Once this data cache is obtained, we will clean it to remove overly large or overly small files, odd file types, duplicates, and other problematic files. We will aim to have 700-1000 viable images for each class after cleaning. Files will be renamed corresponding to animal type and numbered. These cleaned and organized files will be split into groups for training, validation, and testing.

**Objectives:**

This project has two main objectives. The first is to create a data product that achieves the function of identifying an animal type from a photo of said animal and to perform this task with an average accuracy of 92% or higher. When this model analyzes images of animals it has never before seen, from the eight classes it was trained on, it will predict the correct animal type ninety-two or more times out of every one-hundred attempts. The second objective is to complete all steps of developing this model within six weeks from start to delivery.

**Methodology:**

For this project, we will be following the Waterfall methodology model. Waterfall is very effective for small scale projects like this that have a well-defined set of goals and requirements. Progression through the phases is linear and this makes it very easy to gauge progress. Waterfall does not require constant input from the customer, instead only requiring their input at the beginning during requirements gathering, at important milestones, and near the end of the project lifecycle for verification that the deliverable(s) meet the requirements, in good working order, with little to no obvious defects. This methodology has five phases: Requirements, Design, Implementation, Verification, and Maintenance.

In the Requirements phase we will be meeting again with ABC’s website manager, Mr. Jones to further discuss the planning of this project and confirm ABC’s specific expectations and goals. The Design phase is where the product specifications take shape. We will use the requirements gathered from Mr. Jones to determine our path forward and plan the steps needed for our Implementation phase. Implementation is where the bulk of the activity will take place. In this phase, we will follow the plan we laid out in the Design phase to actually code out and build the deliverable. Verification is the process of testing a product to determine if it is ready to go to the customer. Testing of our model will be done internally during the training process and then Mr. Jones will have the opportunity to demo the data product and review our results before signing off on delivery. Deployment and Maintenance is the final step and involves actually deploying the product for use and monitoring for issues to perform necessary maintenance. XYZ’s role in this phase will be limited to providing user instructions at delivery to ABC’s website management team as this product is a single module to be added to ABC’s website. The monitoring and maintenance of the website falls and will continue to fall under Mr. Jones’ purview.

**Funding Requirements:**

The funding required to cover development man hours and use of equipment and facilities for six weeks will be $15,000. This includes all necessary software, data access, administrative and legal fees, and taxes. If an increase to the minimum acceptable accuracy of the model (currently set at 92%+ average) are requested in the requirements phase, we can accommodate these changes, but this will come with additional time and funding requirements attached for which we will provide revised estimates.

**Data Precautions:**

The development of this data product will utilize publicly accessible image data from Kaggle.com and Google image searches. None of this training data is personally identifiable or confidential in any way. Customers who choose to submit photos to ABC Feed Supply once the product is launched will do so of their own volition and ABC would be wise to notify them of how their data will be used in their Terms of Service.

**Developer’s Expertise:**

The developers involved in this project have received their Bachelors’ of Computer Science from respected universities, are versed in several programming languages including Python, Java, SQL, and C++, and have experience working with multiple development platforms. The members of our team have obtained multiple technical certifications related to development and project management. My team members have varying levels and modes of experience and are all industry professionals of the highest degree. Your project is in good hands with us!

Part B

**Project Proposal**

**Problem Statement:**

In today’s marketplace, it is a necessity for a company’s web presence to match or exceed its physical presence. Consumers are more connected to their devices and screens and less connected to their communities and local businesses. Online purchases are continually increasing and in-store, brick-and-mortar purchases are continually declining. Consumers have less time and attention-span than ever before and crave engagement, excitement, and convenience.

ABC Feed Supply has recently expanded the operations of its physical business and its website needs an upgrade, as its current version does not fully address the evolving needs of the customer. While the layout is convenient and simple to use, there is a lack of engaging features that will grab and hold the attention of current and future customers.

**Customer Summary:**

ABC Feed Supply is a mid-size feed producer that sells feeds formulated for eight animal types typically raised as livestock or pets. They have recently expanded their territory from primarily Midwest operations to the Plains states as well by opening more brick-and-mortar outlets. They have a web presence, but would like to add upgraded features to attract new customers and maintain customer relationships. Currently a large portion of their business is in the form of in store purchases, but they also offer online ordering which can be fulfilled with shipping or store pickup. ABC’s customers come from all walks of life, but, regardless of their background, typically fall into two broad categories: farmers and pet owners. The data product we are proposing will help both ABC and its customers. The customers will benefit from an interactive shopping experience that provides engagement and information in an efficient manner and ABC will benefit from the increased website traffic and resulting revenue from purchases as well as the ability to harvest more customer data for use in marketing and planning for future expansions.

**Existing System Analysis:**

ABC currently hosts a pretty basic retail website. There is product content and pricing information available, it has a cart that holds customer items for purchase, customer profiles can hold purchase and shipping data, and it has the ability to process payments. There is a general lack of engaging content or functionality that could increase customer engagement. This could be one reason that the majority of sales are currently in store. The implementation of our data product into the website will increase customer engagement, attract new customers, and as a result sales on the web side will increase.

**Data:**

We will be utilizing publicly available images of cats, chickens, cows, dogs, goats/sheep, horses, pigs, and rabbits for training, validating, and testing the artificial intelligence model and these images will be sourced from Kaggle.com as well as Google image searches. Once this data cache is obtained, we will clean it to remove overly large or overly small files, odd file types, duplicates, and other problematic files. We will aim to have 700-1000 viable images for each class after cleaning. Files will be renamed corresponding to animal type and numbered. These cleaned and organized files will be split into groups for training, validation, and testing. As our data will be running training on a fine-tuned MobileNet model, preprocessing will be done to resize and format images to the MobileNet standard. The preprocessing function will be handled during training and will not permanently affect the original dataset. This will ensure that customers can submit a variety of file size, type, and color scheme to the final product without issue.

**Project Methodology:**

For this project, we will be following the Waterfall methodology model. Waterfall is very effective for small scale projects like this that have a well-defined set of goals and requirements. Progression through the phases is linear and this makes it very easy to gauge progress. Waterfall does not require constant input from the customer, instead only requiring their input at the beginning during requirements gathering, at important milestones, and near the end of the project lifecycle for verification that the deliverable(s) meet the requirements, in good working order, with little to no obvious defects. This methodology has five phases: Requirements, Design, Implementation, Verification, and Maintenance.

In the Requirements phase we will be meeting again with ABC’s website manager, Mr. Jones to further discuss the planning of this project and confirm ABC’s specific expectations and goals. The Design phase is where the product specifications take shape. We will use the requirements gathered from Mr. Jones to determine our path forward and plan the steps needed for our Implementation phase. Implementation is where the bulk of the activity will take place. In this phase, we will follow the plan we laid out in the Design phase to actually code out and build the deliverable. Verification is the process of testing a product to determine if it is ready to go to the customer. Testing of our model will be done internally during the training process and then Mr. Jones will have the opportunity to demo the data product and review our results before signing off on delivery. Deployment and Maintenance is the final step and involves actually deploying the product for use and monitoring for issues to perform necessary maintenance. XYZ’s role in this phase will be limited to providing user instructions at delivery to ABC’s website management team as this product is a single module to be added to ABC’s website. The monitoring and maintenance of the website falls and will continue to fall under Mr. Jones’ purview.

**Project Outcomes:**

The deliverables of this project will include:

* **A fully functional predictive model:** This CNN trained model will take an image as input and predict which of eight classes (cat, chicken, cow, dog, goat/sheep, horse, pig, or rabbit) the image contains.
* **Source code with thorough comment documentation:** The source code used to create this model and all associated functions and commenting used in the development will be turned over to ABC’s website management team for review and reference.
* **Training, Testing, and Validation image datasets:** The image data used in the process of creating the model will also be given to the ABC website management team.
* **A user guide with tips on the use and integration of this feature:** As this proposal is for the design and creation of the image classification model only, there will not be a web friendly user interface included. We will include in the user guide, suggestions for minor code changes to help integrate with ABC’s website.

**Implementation Plan:**

Implementation will depend heavily on the requirements and design phases of our Waterfall methodology and could easily change with a change in scope. Assuming no changes to our current proposal our implementation phases will be:

* Build an image dataset of the eight classes sufficient for training, validation, and testing of the predictive model (aiming for at least 1000 unique images per class).
* Clean the data by removing overly large or small files, abnormal file types, and files that cause errors from the dataset (aiming for at least 700 images per class after removal).
* Separate files into training, validation, and testing folders for each class.
* Download or update all necessary development environs, models, Python, and Python libraries.
* Write code for the model and preprocessing of data as well as associated functions for displaying metric information, image data, and confusion matrices.
* Train and validate the model.
* Test the model against test data, record results into a confusion matrix, and assess results.
* Fine-tune the model or augment the data to improve performance.
* Repeat the train and validate, test, and fine tune steps until acceptable levels of accuracy are achieved.
* Save the trained model.
* Write code for single image predictions and test against the saved model with images not used in previous training, validation, or testing.
* Verify outcomes are similar to the expected accuracy range. If this test fails return to the train and validate step and repeat until acceptable accuracy is achieved.
* Write user guide to include with deliverables.
* Deliver final deliverables to the client for review.

**Evaluation Plan:**

Validation of the model training will be performed during each epoch of the model training function. We will utilize a high verbosity setting during training to receive real-time information, the use of plotting functions to assess accuracy and loss trends during training, and confusion matrices to visualize correct vs. incorrect predictions on test data. Evaluation of the model’s overall success will be based on meeting our target goals for accuracy of 92% and delivery deadline of six weeks from start to finish.

**Resources and Costs:**

The below technical resources will be required in the development of this project:

* Terminals installed with the following:
* Development Environment:
  + JupyterLab 3.4.4
* Operating System:
  + Windows 11
* Processor:
  + Intel Core i3-8130U CPU @ 2.20GHz 2.21 GHz
* Installed RAM:
  + 8.00 GB (7.88 GB usable)
* System type:
  + 64-bit operating system, x64-based processor
* Language and library versions:
  + Anaconda 3
  + Python 3.9.13
  + Numpy 1.21.5
  + TensorFlow 2.11.0
  + Keras 2.11.0
  + Scikit-learn 1.0.2
  + IPython 7.31.1
  + Matplotlib 3.5.2

|  |  |  |
| --- | --- | --- |
| **Resource** | **Description** | **Cost** |
| Above listed technical resources | Terminal with installed development necessities | $0 (already owned by XYZ) |
| Training, Validation, and Test image data | Image datasets used for training, validating, and testing the model | $0  (Data can be accessed free of charge, via Kaggle.com or open-source web search like Google image search) |
| Development team time and effort | Humans that will build and train the AI solution over a six-week time period | $15,000 |
|  | **Total** | $15,000 |

**Timeline and Milestones:**

|  |  |  |
| --- | --- | --- |
| Pre-development | Gather requirements. | Acceptance of proposal |
| Week 1 | Verify requirements.  Design solution.  Obtain image data.  Clean and organize data.  Partition data. |  |
| Week 2 | Download or update all necessary technical material.  Write code for the training model, image preprocessing functions, and functions for performance visualization to design specifications. |  |
| Week 3 | Train, validate, test, and evaluate model.  Fine -tune training model and augment data as necessary to improve accuracy. |  |
| Week 4 | Train, validate, test, and evaluate model.  Fine -tune training model and augment data as necessary to improve accuracy. |  |
| Week 5 | Train, validate, test, and evaluate model.  Fine -tune training model and augment data as necessary to improve accuracy. |  |
| Week 6 | Write code for single image predictions.  Verify accuracy.  Finalize testing.  Write user guide. | Deliver completed data product and associated literature to the ABC Feed Supply website management team. |

Part C

**Application**

The C964 Project Data folder and contents I will be submitting are laid out as such:

C964 Project Data

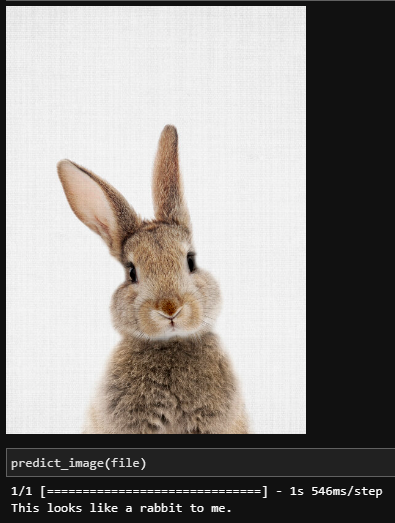
* .ipnyb\_checkpoints
  + This folder contains the checkpoints for the below .ipnyb files
* Barnyard Images
  + This folder contains all images that were not removed for nonconformity issues. The folders named after each class of animal are the sources for the train, valid, and test folders and contain the excess files that were not used in training, validation, or testing. These files can be ignored or used for single prediction testing. The test, train, and valid folders were used in training, validation, and testing. The folder titled verification should be used for single prediction testing/demoing and currently has one picture from each animal class. Additional single prediction testing images can be added to this folder if desired either from the excess images or sourced elsewhere.
* Models
  + This folder contains the save file for our fully trained model, which achieved %99.54 training accuracy, %93.75 validation accuracy, and %94.25 prediction accuracy on a set of 400 test images.
* BarnyardAnimalClassifierForTraining.ipynb
  + This file contains all the building, training, validation, and testing code for our model and multiple graphics including plots of loss and accuracy during training as well as sample images and a confusion matrix of the test set predictions. Each block of code is commented for clarity as to its functionality.
* BarnyardAnimalClassifierForDemo.ipynb
  + This file is a stripped-down version of the BarnyardAnimalClassifierForTraining.ipynb file with the model building and training sections removed. It is ideal for a demonstration of functionality as it loads the saved model without showing all the build data. The functionality allows for demos of the test phase with accompanying confusion matrix and the ability to upload an image for single prediction. Each block of code is commented for clarity as to its functionality.
* BarnyardAnimalClassifierForABCWebTeam.ipynb
  + This file is what would be submitted to ABC Feed Supply as a deliverable. It only contains the necessary functionality of loading the trained model, displaying an image, and making a prediction on that image. Each block of code is commented for clarity as to its functionality.
* FileManipulation.ipynb
  + This file contains the functions used for cleaning our dataset, renaming all files to our naming convention, and splitting out and populating the valid, test, and train folders. Each block of code is commented for clarity as to its functionality.

Part D

**Post-Implementation Report**

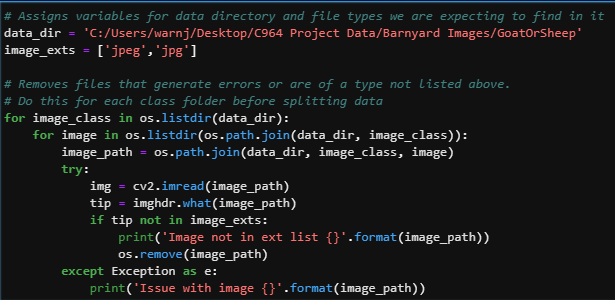
**Project Purpose:**

The ABC Feed Company has requested a feature that can suggest feed products to a customer when they upload a photo of their animal to ABC’s website. ABC produces and supplies feed products for eight types of animals: cats, dogs, chickens, cows, goats/sheep, horses, pigs, and rabbits. In response to this request, I have developed a deep learning CNN fine-tuned from the popular MobileNet model that can predict which of the 8 classes of animal a photo depicts with an accuracy of %94.25 across a test set of 400 images. I have added the functionality to predict on a single image and this continues to function at a high rate of accuracy on images never before seen by the model. ABC’s website management team has integrated this feature functionality into their website design.

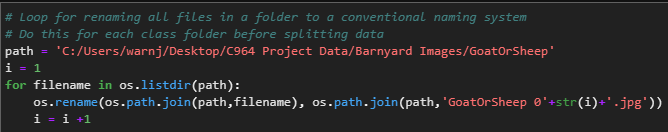


**Datasets:**

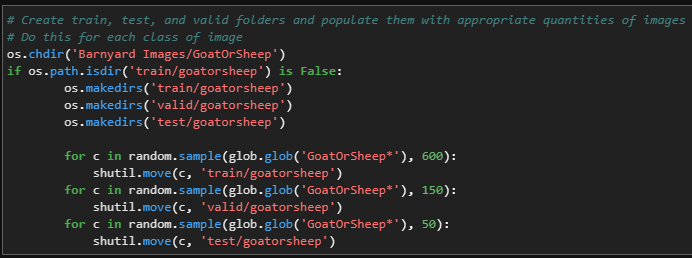
I utilized publicly available images of cats, chickens, cows, dogs, goats/sheep, horses, pigs, and rabbits for training, validating, and testing the artificial intelligence model and these images were sourced from Kaggle.com as well as Google image searches. Once this data cache was obtained, I cleaned it to remove overly large or overly small files, odd file types, duplicates, and other problematic files. The removal of large, small, and duplicate files was completed by manually sorting and removing via my machine’s file manager. Files were renamed corresponding to animal type and numbered sequentially via a function ran in JupyterLab to manipulate the files. These cleaned and organized files were split randomly into groups for training, validation, and testing which was also carried out via a function in JupyterLab. As this data ran training on a fine-tuned MobileNet model, preprocessing was done to resize and format images to the MobileNet standard. The preprocessing function was handled during training and did not permanently affect the original dataset. This ensures that customers can submit a variety of file size, type, and color scheme to the final product without issue. All images that were not removed from the dataset in the cleaning phase, including testing, validation, training, and unused images are being submitted within the project folder for review.



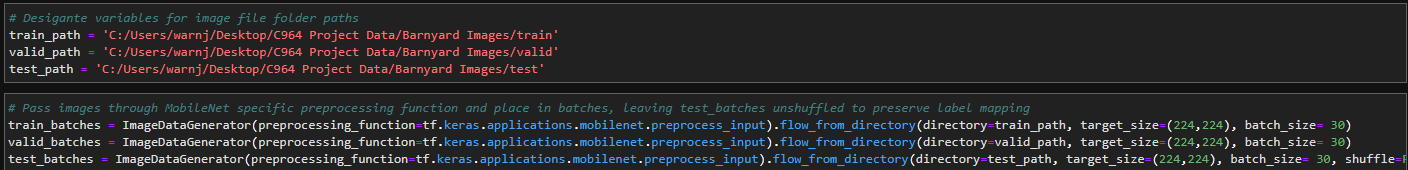
Function for cleaning image files.



Function for renaming files.



Function for splitting files into train, test, and valid groupings.

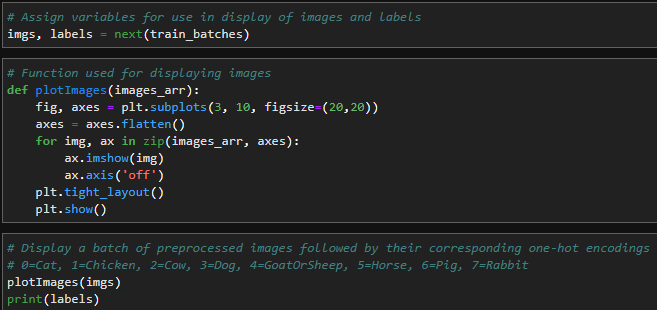


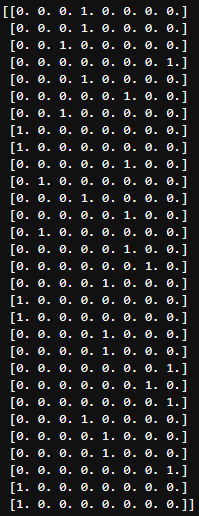
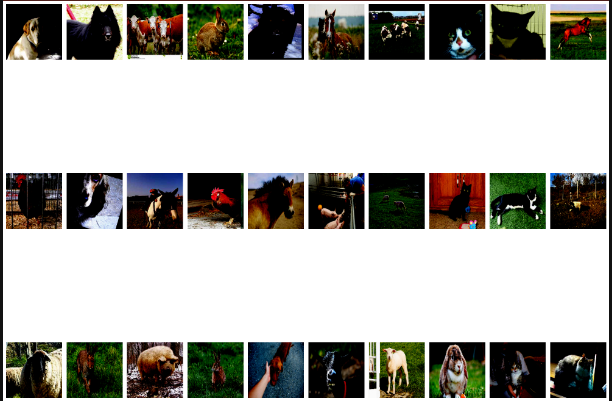
Code used to preprocess images to MobileNet standard. The test files were left unshuffled to maintain label mapping for confusion matrix plotting purposes.

**Data Product Code:**

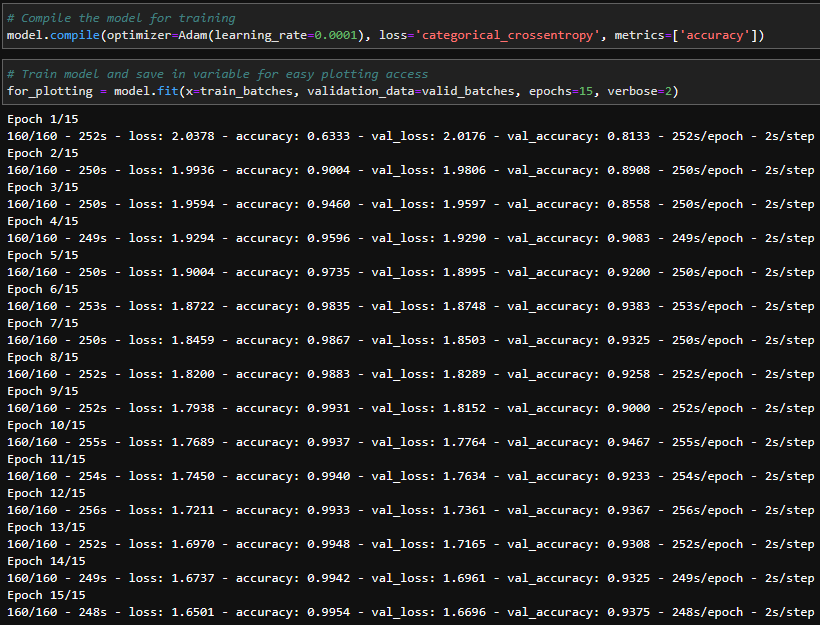
Building and training the predictive model involved several steps and a lot of repetition. I actually built dozens of models and fine-tuned the most successful attempts following a linear production path and iterated through the training, validation, testing, and assessment phases until my accuracy metrics were acceptable. The order of steps for the final model build training was:

1. Import dependencies.
2. Download (first time) or load the MobileNet model.
3. Display the model to study architecture and trainable/non-trainable parameters.
4. Designate variables for train, valid, and test paths and preprocess the images.
5. Replaces the output layer of the MobileNet model with an 8-neuron output layer.
6. Save this as a new model.
7. Freeze all but the last 18 layers of the new model as non-trainable
8. Display and compare the architecture and parameter trainability to MobileNet.
9. Write a function for and then display a batch of training images with their one-hot encoded labels.

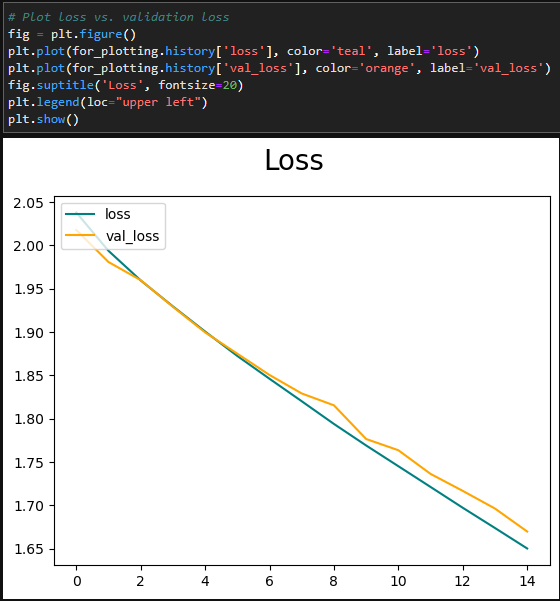


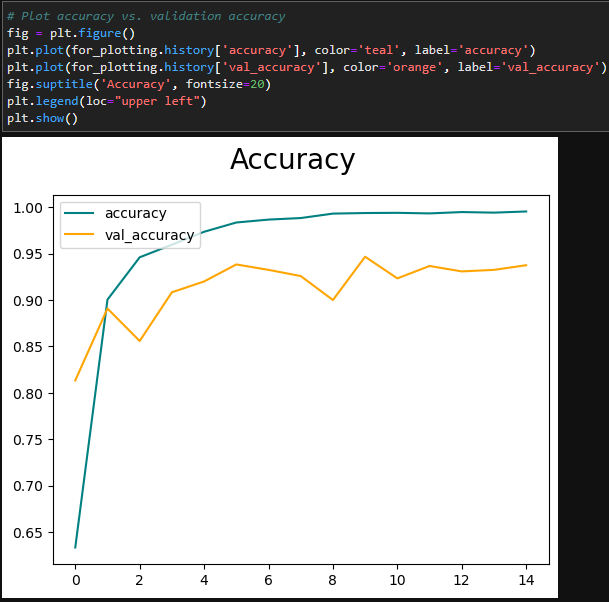


1. Compile the model for training using categorical cross entropy to measure loss and setting metrics to accuracy.
2. Train the model and save the training history for plotting purposes.

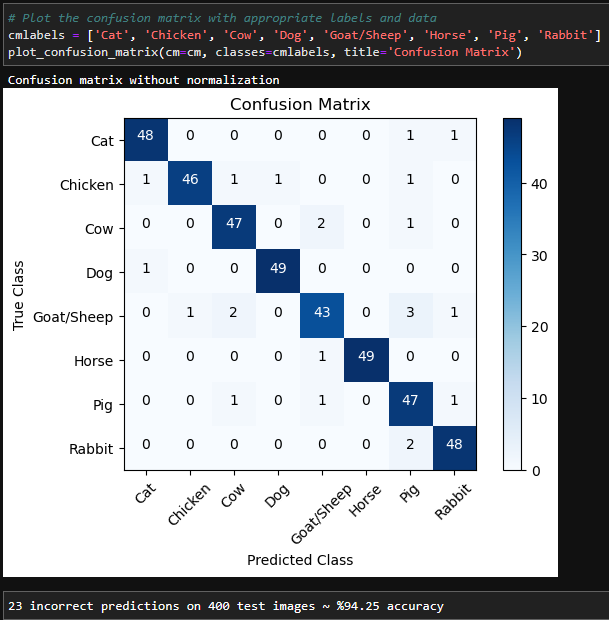


1. Plot and assess loss vs. validation loss and accuracy vs. validation accuracy.





1. Run predictions on the test data set of images and save to a variable for plotting.
2. Define, plot, evaluate, and calculate accuracy of a confusion matrix with test data.



1. Save the model for future use.
2. Define functions for loading, displaying, and preprocessing single images.
3. Load, display, preprocess, and make predictions on a series of single images.



**Objective Verification:**

My main objectives when completing this project were:

* To have a working predictive CNN which achieved at least %92 accuracy on test data.
  + This model achieved %94.5 accuracy on a test set of 400 images.
* To finish the project in less than six weeks which coincided with ABC’s needs in the scenario and also with my end of term for graduation.
  + It has been roughly five weeks since I started this course, and I will be submitting this tonight for my first attempt with time for revisions.

These objectives have been met successfully.

**Effective Visualization and Reporting:**

I used several descriptive methods (many pictured above) in the course of developing this project from displaying images to plotting structures. I am a big fan of using print statements and other visuals for confirmation of correct operation during the development process. I display images to verify their file paths and the functions that call them are working correctly. The loss and accuracy graphs helped me visualize the progress of my predictive CNN models while working on built-from-scratch models that tested in the forty to fifty percent range up to my final model that hit over %94. The accuracy and loss graphs really helped to point out issues like over or underfitting or the dreaded vanishing or exploding gradients which can point the developer in the right direction when deciding things like learning rate, number of training epochs, etc. Confusion matrices allowed me to see if certain classes were training better than others in order to make decisions about any data augmentation that may have needed to be performed.

**Accuracy Analysis:**

As a specific accuracy threshold was one of my main objectives, I have discussed accuracy in nearly every section of this report and there is a training vs. validation accuracy graph as well as a screenshot of the training progress through epochs showing the accuracy and loss measures during each stage of training for my final model. These are displayed in the Data Product Code section on pages 22 and 23.

**Application Testing:**

After each CNN predictive model was trained, I ran the model.predict method which sent my 400 test images through the predictive model and measured for accuracy. These results were plotted to confusion matrices, and I assessed the results. If accuracy was lacking, I used the information garnered from assessment to adjust the model (fine-tuning) or data (data augmentation) and then trained a new model. The cycle of training, testing, evaluating, and adjusting went on for several days or weeks before arriving at my %94+ accuracy model.

**Application Files:**

The C964 Project Data folder and contents I will be submitting are laid out as such:

C964 Project Data

* .ipnyb\_checkpoints
  + This folder contains the checkpoints for the below .ipnyb files
* Barnyard Images
  + This folder contains all images that were not removed for nonconformity issues. The folders named after each class of animal are the sources for the train, valid, and test folders and contain the excess files that were not used in training, validation, or testing. These files can be ignored or used for single prediction testing. The test, train, and valid folders were used in training, validation, and testing. The folder titled verification should be used for single prediction testing/demoing and currently has one picture from each animal class. Additional single prediction testing images can be added to this folder if desired either from the excess images or sourced elsewhere.
* Models
  + This folder contains the save file for our fully trained model, Models/BarnyardAnimalClassifierFinal.h5, which achieved %99.54 training accuracy, %93.75 validation accuracy, and %94.25 prediction accuracy on a set of 400 test images.
* BarnyardAnimalClassifierForTraining.ipynb
  + This file contains all the building, training, validation, and testing code for our model and multiple graphics including plots of loss and accuracy during training as well as sample images and a confusion matrix of the test set predictions. Each block of code is commented for clarity as to its functionality.
* BarnyardAnimalClassifierForDemo.ipynb
  + This file is a stripped-down version of the BarnyardAnimalClassifierForTraining.ipynb file with the model building and training sections removed. It is ideal for a demonstration of functionality as it loads the saved model without showing all the build data. The functionality allows for demos of the test phase with accompanying confusion matrix and the ability to upload an image for single prediction. Each block of code is commented for clarity as to its functionality.
* BarnyardAnimalClassifierForABCWebTeam.ipynb
  + This file is what would be submitted to ABC Feed Supply as a deliverable. It only contains the necessary functionality of loading the trained model, displaying an image, and making a prediction on that image. Each block of code is commented for clarity as to its functionality.
* FileManipulation.ipynb
  + This file contains the functions used for cleaning our dataset, renaming all files to our naming convention, and splitting out and populating the valid, test, and train folders. Each block of code is commented for clarity as to its functionality.

The IDE and libraries used to develop the final solution are:

* Development Environment:
  + JupyterLab 3.4.4
* Language and library versions:
  + Anaconda 3
  + Python 3.9.13
  + Numpy 1.21.5
  + TensorFlow 2.11.0
  + Keras 2.11.0
  + Scikit-learn 1.0.2
  + IPython 7.31.1
  + Matplotlib 3.5.2

When running BarnyardAnimalClassifierForABCWebTeam.ipynb (the final data product) only the below are required:

* Models/BarnyardAnimalClassifierFinal.h5
* BarnyardAnimalClassifierForABCWebTeam.ipynb
* JupyterLab 3.4.4
* Python 3.9.13
* Numpy 1.21.5
* TensorFlow 2.11.0
* Keras 2.11.0
* IPython 7.31.1
* At least one image to run prediction on

**User Guide:**

1. Download and extract the project folder C964 Project Data.
2. Download/update JupyterLab, Python, Numpy, TensorFlow, Keras, and IPython.
3. Open JupyterLab.
4. Open the BarnyardAnimalClassifierForABCWebTeam.ipynb file.
5. Edit the filepath in red type in the code block that begins with def prepare\_image(file) to match the image file location on your machine after download. Pictured below.

Text

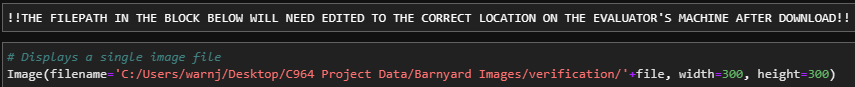
Description automatically generated

1. Choose an image for prediction from, or move one to, the verification folder.
2. Edit file= to indicate the filename of the chosen image

Graphical user interface

Description automatically generated

1. Edit Image(filename= to indicate the filepath of the chosen image to match the location after download to your machine.



1. Tap the Restart Kernel and Run button in JupyterLab.

Graphical user interface, text, application, chat or text message

Description automatically generated

1. View results.
2. Be amazed.
3. Repeat as many times and with as many images as necessary.

**Summation of Learning Experience:**

This project is the culmination of the last two years of education. In the programming courses of this degree program, I was exposed to working with multiple languages including Python which I chose to use for this project. Project Management and Software Engineering courses prepared me for the proposal portions of this project by exposing me to design and implementation methodologies. I also learned a lot about CNNs, machine learning, and image classification for my Artificial Intelligence Task 3 proposal which was knowledge that translated well to this project.

For this project I also learned from sources outside of WGU. I utilized YouTube video tutorial series from Mandy @ DeepLizard that were very helpful in understanding the coding, terminology, and strategies of improvement for the CNNs that I ended up training. I also utilized the API documentation for the add on python libraries when unsure how certain functions operated. Finally, all of the image data was sourced from Kaggle.com from Google image search if insufficient data could be found on Kaggle. Most of the goat, pig, and rabbit images came from Google whereas, the dog, cat, cow, horse, sheep, and chicken images were sourced from Kaggle.

I’ve always believed that we as humans should constantly seek knowledge. The realm of IT is constantly evolving and the technology of today will be obsolete in a year it seems, so lifetime learning is a key factor especially if you want to be successful in this industry. Honestly, I found much of my degree program less than compelling, but this final project where everything came together in the programming portion was very rewarding, even if the accompanying 6000-word, 30-page paper was torture! I look forward to learning and trying new things while expanding my portfolio after graduation and exposing myself to new experiences in my new career path.

**References**

[Keras with TensorFlow Course - Python Deep Learning and Neural Networks for Beginners Tutorial - YouTube](https://www.youtube.com/watch?v=qFJeN9V1ZsI&list=PL1K2uRHNx-Rb-Clcj-kLZNkNpw6vdalJ1&index=22)

[Cat and Dog | Kaggle](https://www.kaggle.com/datasets/tongpython/cat-and-dog)

[Dog vs Not-Dog | Kaggle](https://www.kaggle.com/datasets/danielshanbalico/dog-vs-not-dog)

[Animal Species Classification - V3 | Kaggle](https://www.kaggle.com/datasets/utkarshsaxenadn/animal-image-classification-dataset)

[Animals-10 | Kaggle](https://www.kaggle.com/datasets/alessiocorrado99/animals10)

[Animal Images 5 Classes Bird Cat Dog Fish Rabbit | Kaggle](https://www.kaggle.com/datasets/ashishpahal/animal-images-5-classes-bird-cat-dog-fish-rabbit)

[Mammals classification | Kaggle](https://www.kaggle.com/datasets/anirudhg15/mammals-classification)

[cat vs rabbit | Kaggle](https://www.kaggle.com/datasets/muniryadi/cat-vs-rabbit)