**Title and Team Members**

Animal Recognition – Classification by Using Features

Team Members: Asil Fawalha, Waiyat Hamdani

**Responsibility of Each Team Member**

Asil Responsibility:

* Importing dataset
* Extract features from the dataset
* Splitting the dataset (half of it)
* Couple things on the training part.
* Classification report of classifier
* Decision tree visualization

Waiayt Responsibility:

* Splitting the dataset (half of it)
* Training decision tree classifier
* Testing accuracy
* Confusion matrix

**Introduction:**

* **Problem statement**

Animal recognition technology is a process or classification technique that identifies animal faces in artificial intelligence. The project is basically about the animal classification by features. Classification can be done either using images or a set of features like hair color, height, weight, teeth, skeleton etc. For this project classified animals based upon dataset that contains features. So, the classification is done by using these features. In this experiment, we tested decision tree algorithm to automatically identify the type of each animal by extracting features. We defined a group of 16 features that offered the best resulting accuracy in classifying the animal kind.

* **Objective**

Our main goal in this project is to classify the animal from any given dataset, and surely based on features. This project will help the big zoo to put all the same animal in the right cage.

* **Motivation**

We are interested in the process of animal recognition and how an artificial intelligence can detect and distinguish between different animals. We wanted to use this technology to try and detect animal from any given dataset that contains features. Which varies greatly different from one animal to another. For example, lion and puma are in the same type of class, so we wish to see how we could train a machine to distinguish between 101 animals.

**Related Work**

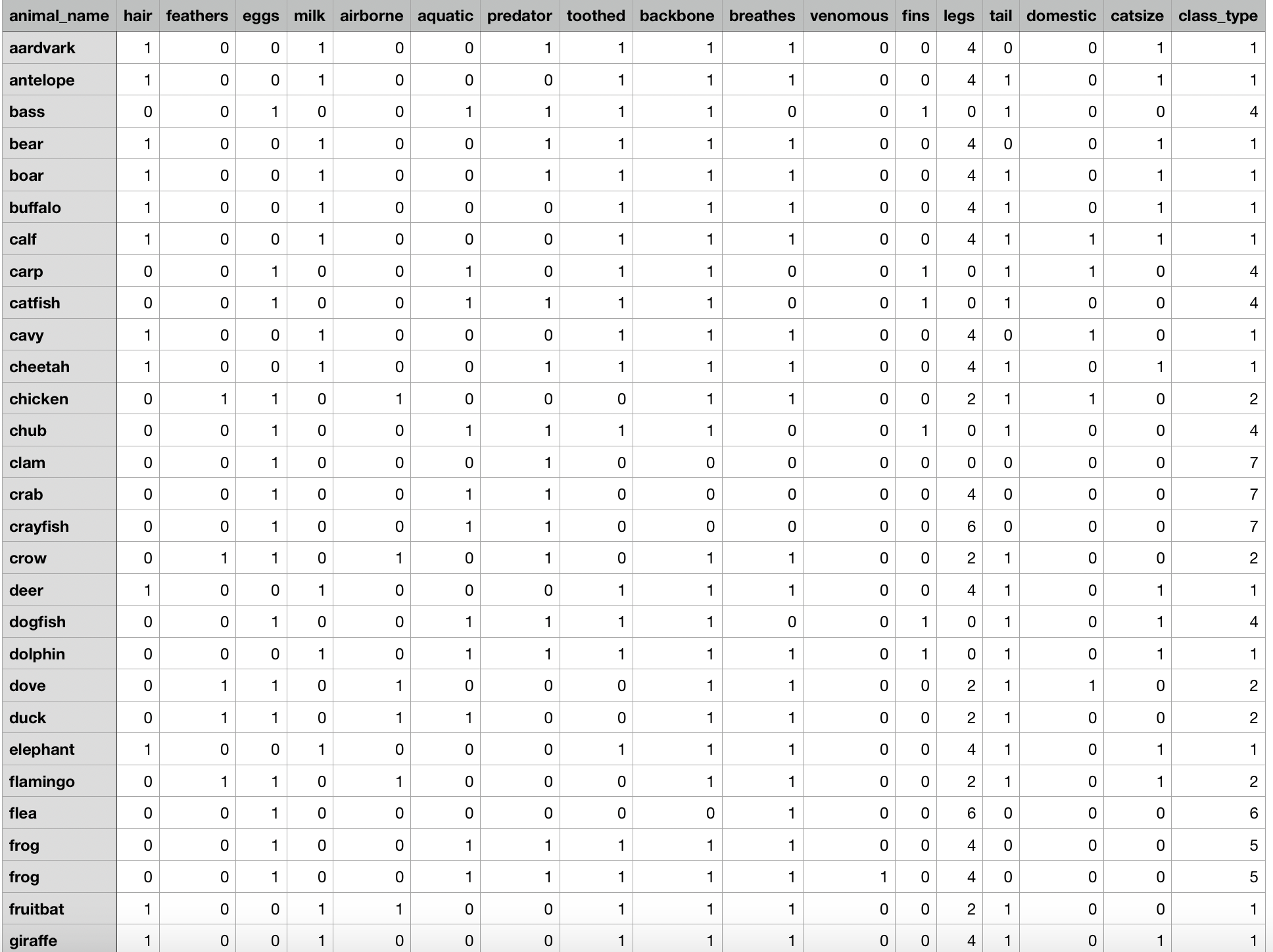
There are so many researches that have been in the field of animal recognition by machine learning. Before we start that project, we though that the best machine in animal classification will be Convolutional Neural Network, but we were wrong on some points. After we started the project, we changed our opinion on using CNN for some reasons. One of these reasons is the better ability that we have in using decision tree algorithm, since we already have an experience in using decision tree. So, this project is done by using decision tree algorithm.

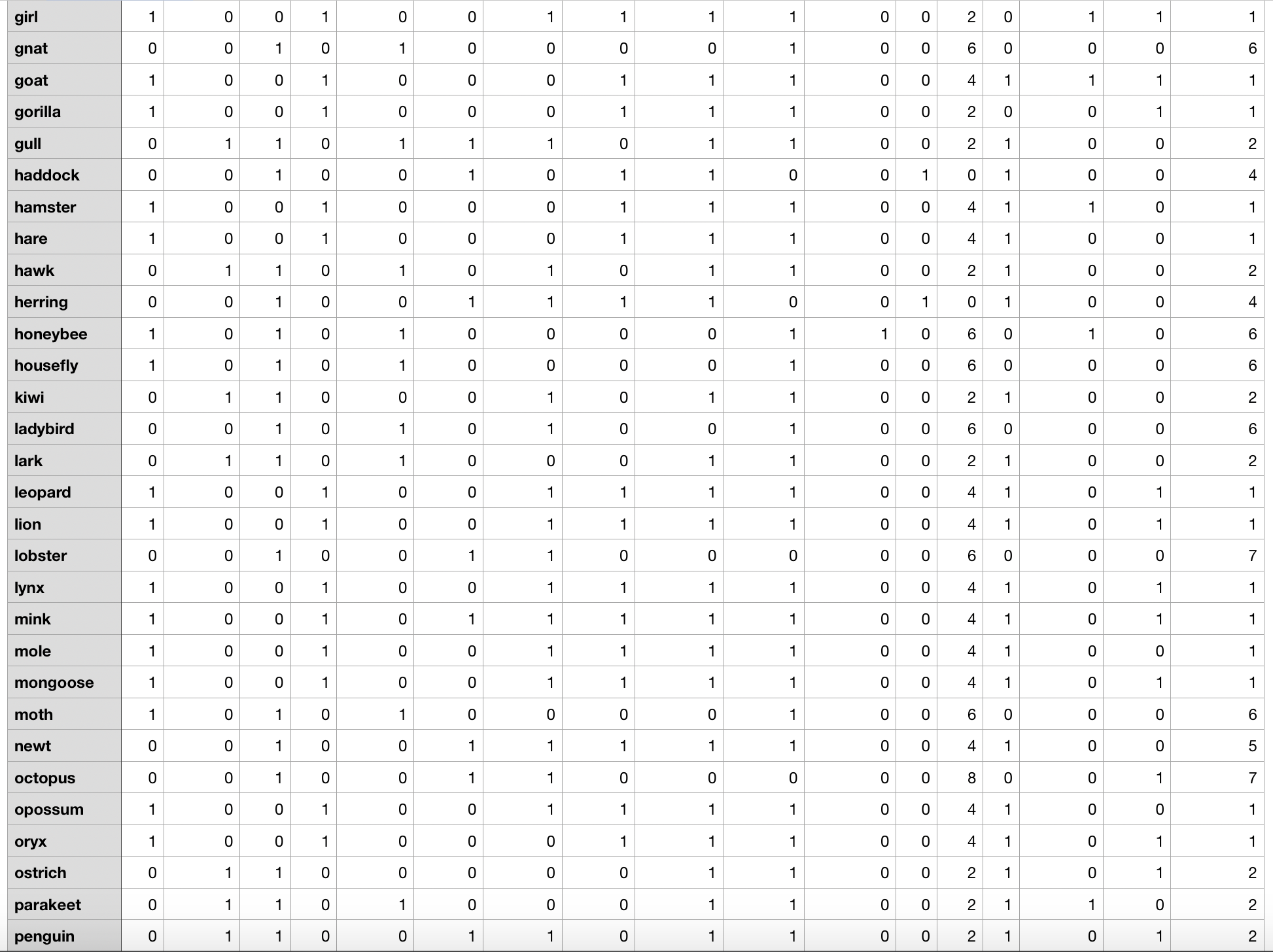
**Data**

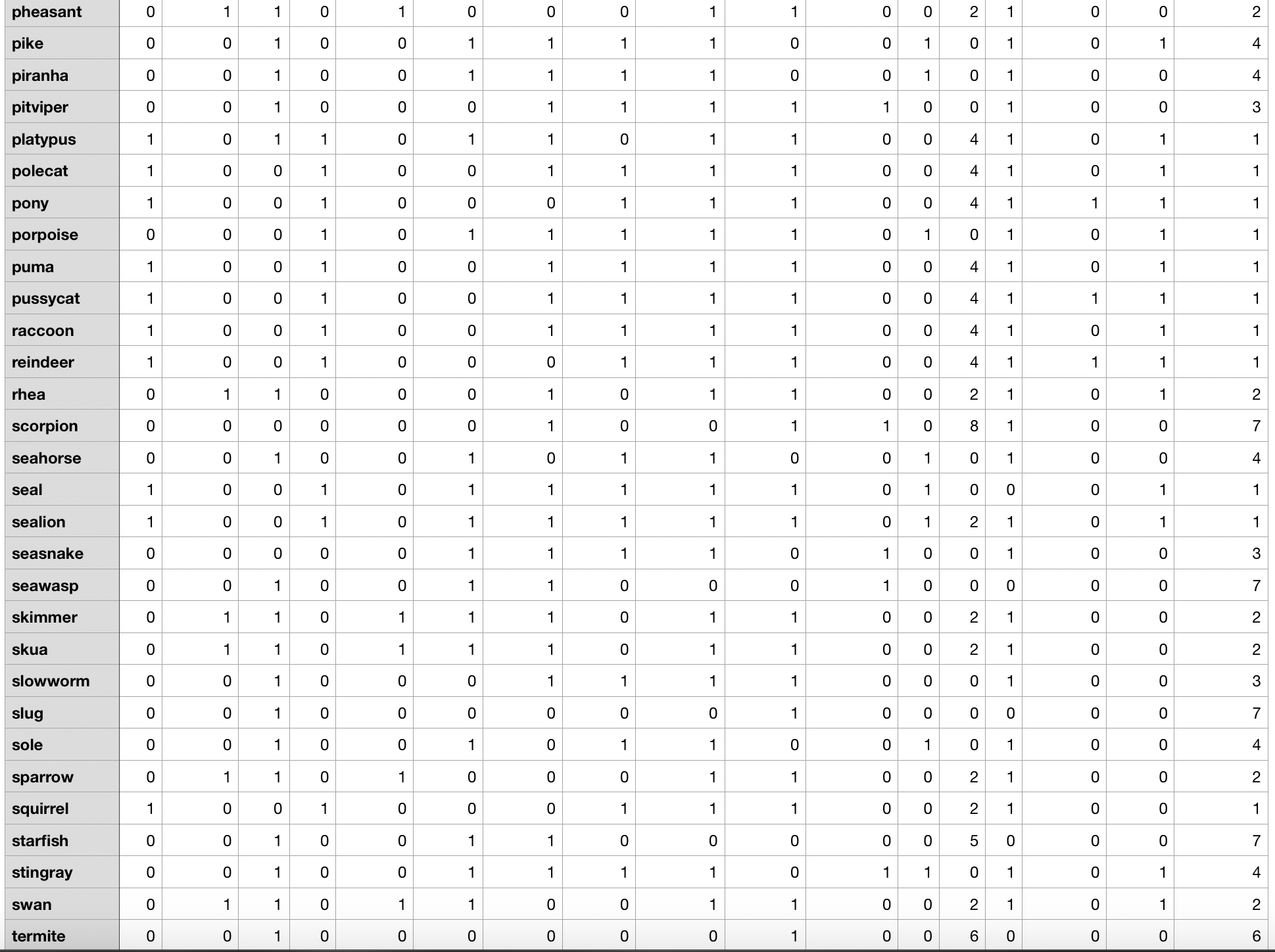
* **Data Source and Format**

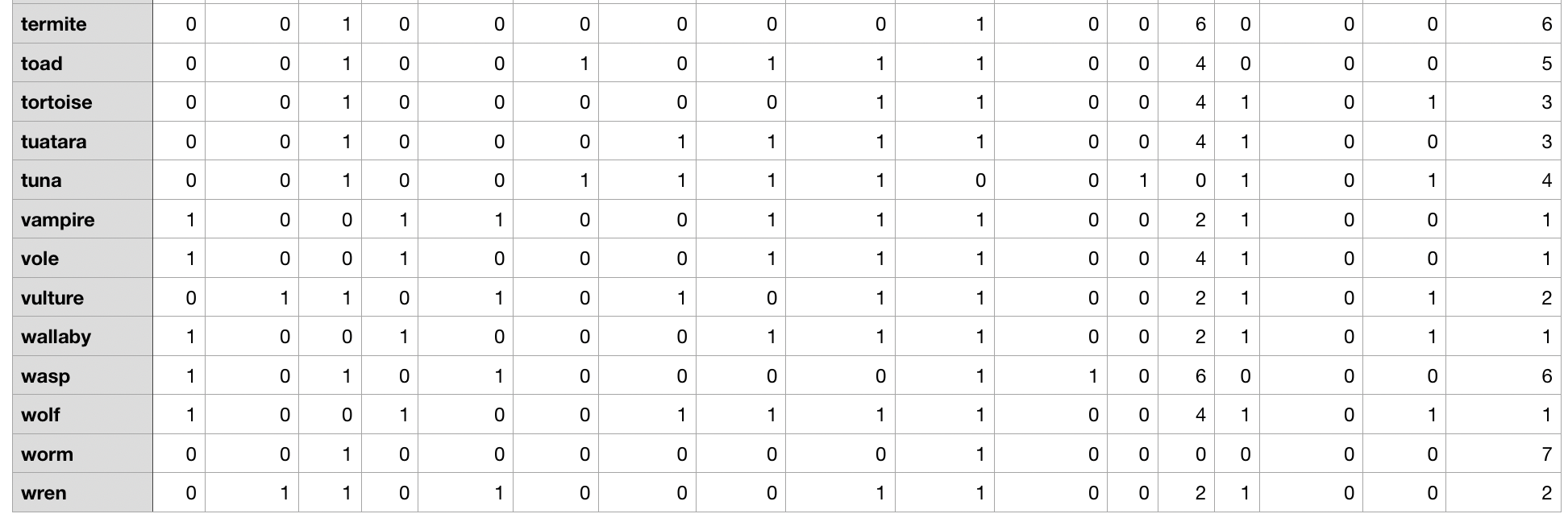
For this project we have a dataset consists of 101 animals from a zoo. The purpose for this dataset is to be able to predict the classification of the animals, based on the variables (16 features). And we have realized that this dataset is a good one for animal recognition using Decision Tree algorithm. The dataset that we are worked on is shown below:

* + Zoo.csv (18 columns):
    - This dataset consists of 101 animals from a zoo.
    - The purpose for this dataset is to be able to predict the classification of the animals, based on the variables.









* **Data Example**

Dataset example for animals: <https://www.kaggle.com/uciml/zoo-animal-classification>

**Features**

* Hair
* Feathers
* Eggs
* Milk
* Airbone
* Aquatic
* Predator
* Toothed
* Backbone
* Breaths
* Venomous
* Fins
* Legs
* Tailk
* Domestic
* Catsize

**Methodology**

* **Preprocessing- Stratified Shuffle Split**

Utilizing mainly python, sklearn, pandas, numpy, and Decision Tree algorithm. In order to start recognizing animal based on features. First, we imported the dataset “zoo.csv”. Second, we extracted a list of features from the imported dataset. For example, hair, feathers, eggs, milk, airborne, aquatic, etc. Third, we worked on the splitting of the dataset to see the distribution of the classes, but the dataset is not distributed equally for each class. So, we did stratified splitting (to keep the distribution equal for both test and train parts) of dataset into testing and training parts. Then we got the distribution in a test set.

* **Features- A List of Features**

In choosing our features, we looked at what features provided the best results in another related research. However, we didn’t choose all the features used in other studies. We were looking for features who has released high accuracy.

Here are the 16 features that we used in the animal classification experiment:

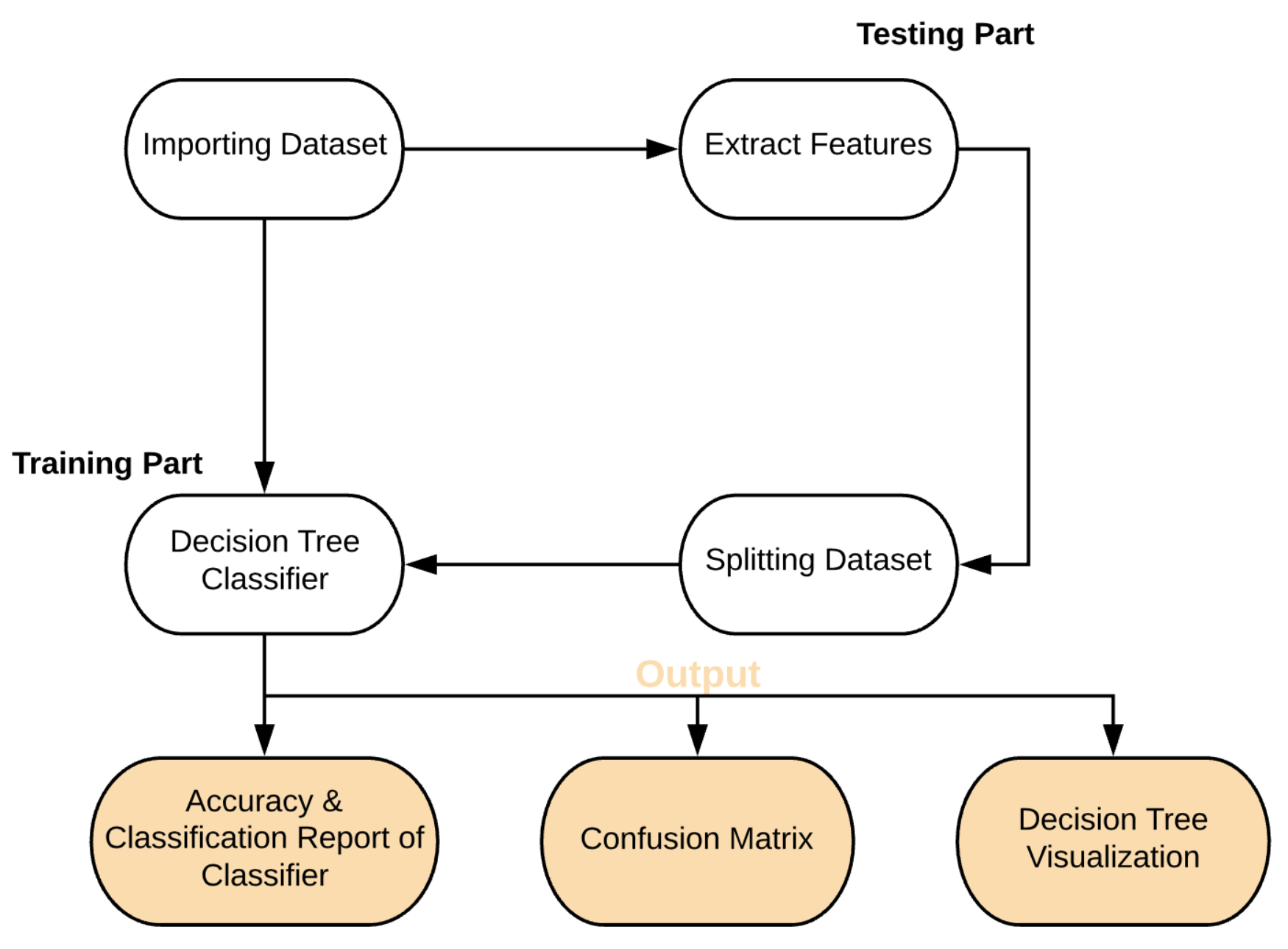
|  |  |
| --- | --- |
| Hair | Backbone |
| Feathers | Breaths |
| Eggs | Venomous |
| Milk | Fins |
| Airborne | Legs |
| Aquatic | Tail |
| Predator | Domestic |
| Toothed | Catsize |

* **Classifier- Decision Tree**

The classification is done by using a classifier known as “Decision Tree”. After we finished the pre-processing, then we applied Decision Tree to the data.

* **Design of Experiment- Data Division (Training/Testing)**

This section is showing a zoom out picture of our experiment design:



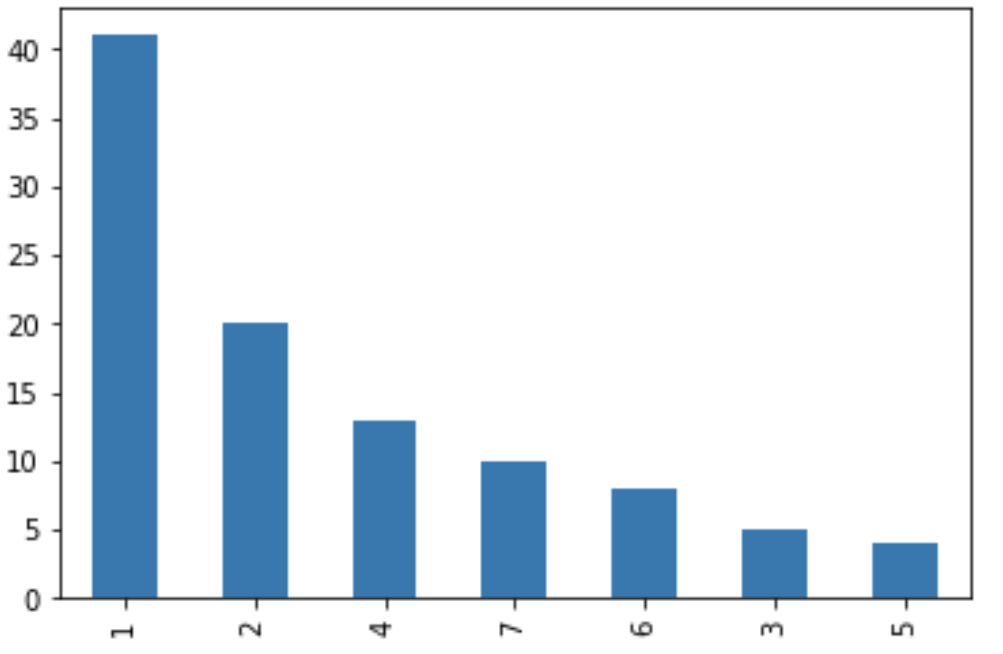
* **Evaluation Matrics**

In addition to find the accuracy of the model, we extracted some additional statistics that detail the performance of the model. We calculated the precision, recall, f1-score, support, macro average, and weighted average. Precision, in short, represents the number of true positive divided by the number of true positives plus the number of false positives. Recall represents the number of correct results divided by the number of results that should have been returned. F1-score represents the harmonic mean of the precision and recall, where an f1-score reaches its best value at 1 (perfect precision and recall). Support represents the number of samples of the true response that lie in the class.

* **Results (Tables/graphs)**
  + The table below represents the class precision, recall, f1-score, support, macro average, weighted average, and the testing accuracy from the decision tree classifier:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Precision | Recall | F1-score | Support |
| 1 | 1.00 | 1.00 | 1.00 | 13 |
| 2 | 1.00 | 1.00 | 1.00 | 6 |
| 3 | 1.00 | 0.50 | 0.67 | 2 |
| 4 | 1.00 | 1.00 | 1.00 | 4 |
| 5 | 1.00 | 1.00 | 1.00 | 1 |
| 6 | 1.00 | 1.00 | 1.00 | 2 |
| 7 | 0.75 | 1.00 | 0.86 | 3 |
| Accuracy |  |  | 0.97 | 31 |
| Macro Avg. | 0.96 | 0.93 | 0.93 | 31 |
| Weighted Avg. | 0.98 | 0.97 | 0.97 | 31 |
| Experiment Testing Accuracy: 0.967741935483871 | | | | |

* + The bar chart below represents the distribution of classes. Dataset is not distributed equally for each class:



* **class 1: Mammals (41)**

aardvark, antelope, bear, boar, buffalo, calf, cavy, cheetah, deer, dolphin, elephant, fruitbat, giraffe, girl, goat, gorilla, hamster, hare, leopard, lion, lynx, mink, mole, mongoose, opossum, oryx, platypus, polecat, pony, porpoise, puma, pussycat, raccoon, reindeer, seal, sealion, squirrel, vampire, vole, wallaby, wolf.

* **class 2: Birds (20)**

chicken, crow, dove, duck, flamingo, gull, hawk, kiwi, lark, ostrich, parakeet, penguin, pheasant, rhea, skimmer, skua, sparrow, swan, vulture, wren.

* **class 3: Reptiles (5)**

pitviper, seasnake, slowworm, tortoise, tuatara.

* **class 4: Fish (13)**

bass, carp, catfisht, chub, dogfish, haddock, herring, pike, piranha, seahorse, sole, stingray, tuna.

* **class 5: Amphibian (4)**

frog, frog, newt, toad.

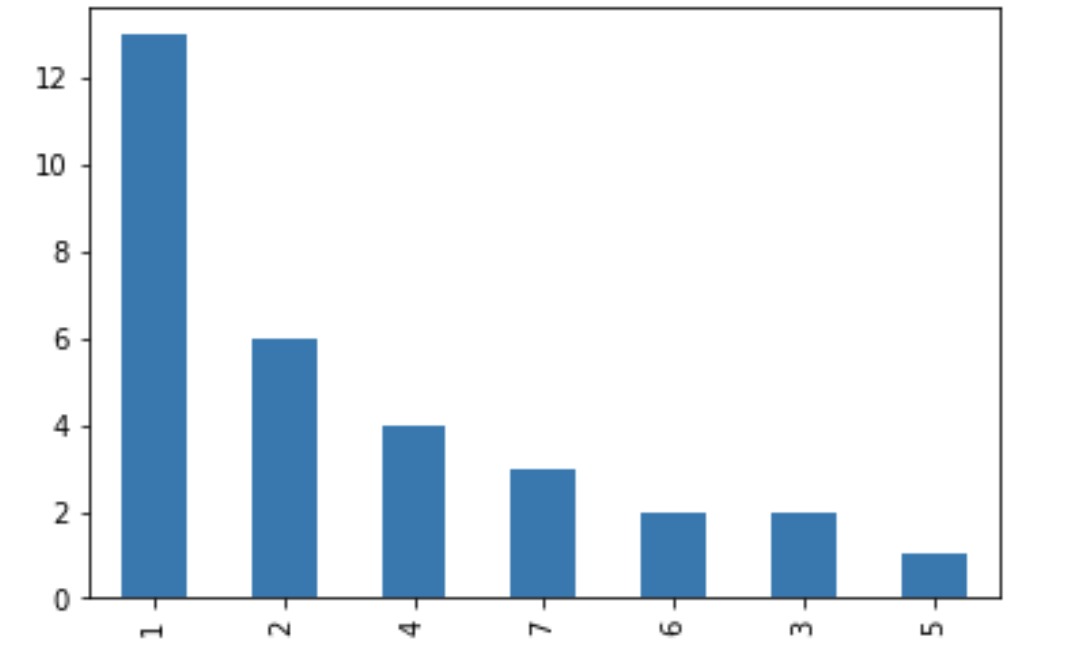
* **class 6: Bugs (8)**

flea, gnat, honeybee, housefly, ladybird, moth, termite, wasp.

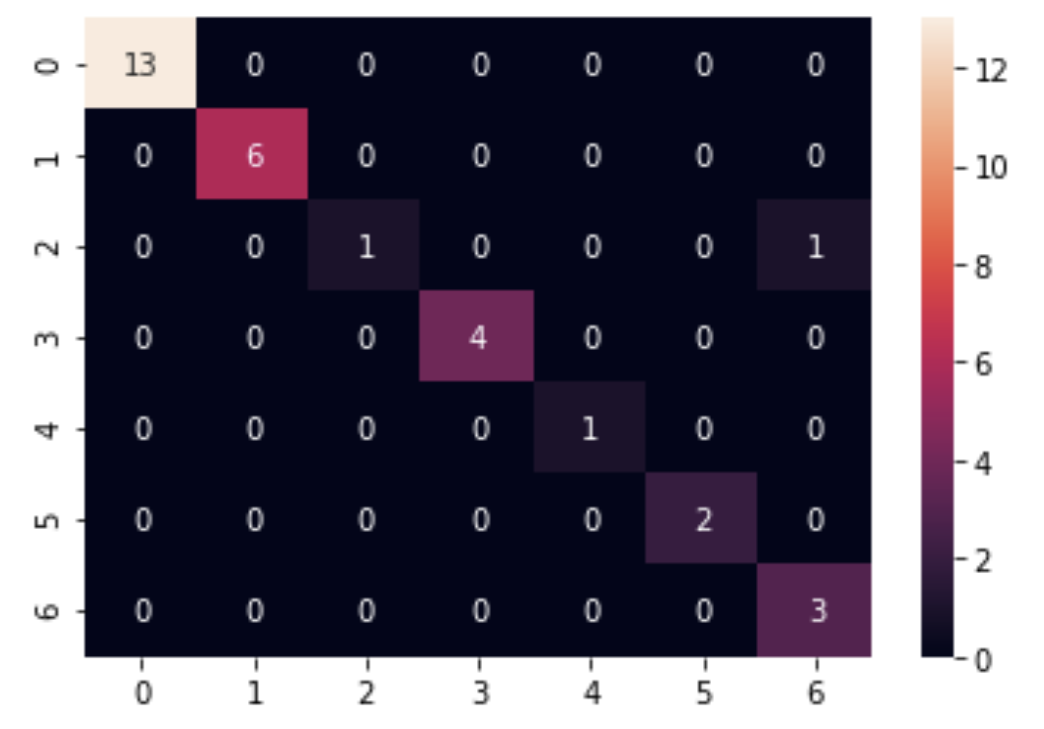
* **class 7: Invertebrate (10)**

clam, crab, crayfish, lobster, octopus, scorpion, seawasp, slug, starfish, worm.

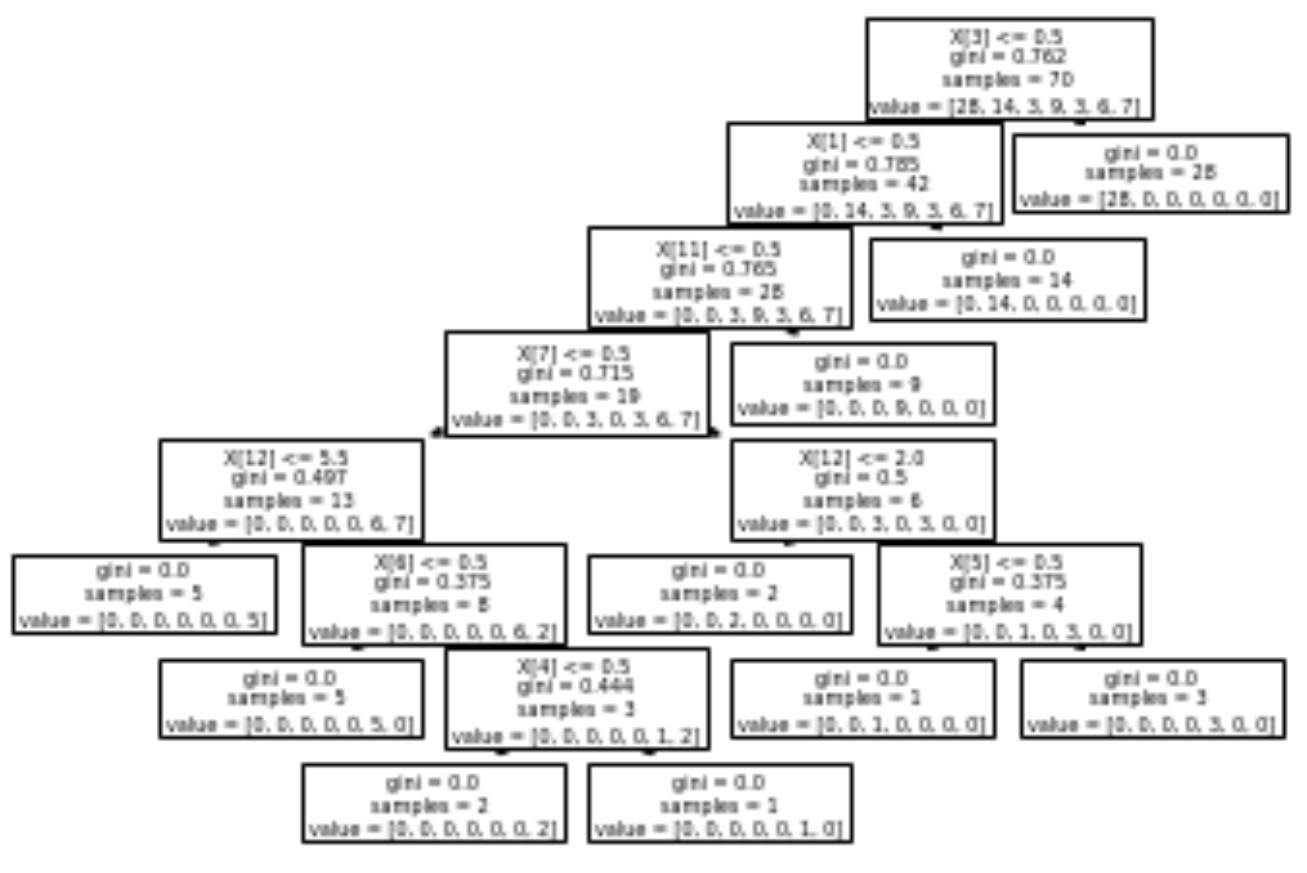
* + The bar chart below represents the distribution of classes in a test set of 31 animals:



* Class 1 Mammal: 13
* Class 2 Bird: 6
* Class 3 Reptile: 2
* Class 4 Fish: 4
* Class 5 Amphibian: 1
* Class 6 Bug: 2
* Class 7 Invertebrate: 3
  + The confusion matrix of the experiment:



* + The decision tree visualization of the experiment:



* **Analysis of the Results**

Since our main goal for this project is to classify the animal from any given dataset based on features and to help the big zoo to put the same animal in the right place. So, after we applied decision tree algorithm to the data, we tested the dataset with the 16 features, 71 animals for training and 31 animals for testing. We came up with a breakdown of which animals are in which type. We came up with 7 classes, each class has a group of animals that could live in the same place. Class 1 group is for mammal, class 2 group is for bird, class 3 group is for reptile, class 4 group is for fish, class 5 group is for amphibian, class 6 group is for bug, and class 7 group is for invertebrate.

**Conclusion**

* **Discuss Any Limitation**

There was no limitation in this project since it was all up to us as a team. we had the option to choose any project and to use any algorithm based on references work.

* **Discuss Any Issues Not Resolved**

In this experiment, we defined 16 features to increase the accuracy of animal classification. We achieved an accuracy of 0.97 when these 16 features were combined with decision tree classifier. in short, everything we planned to, was successfully done in the right way.

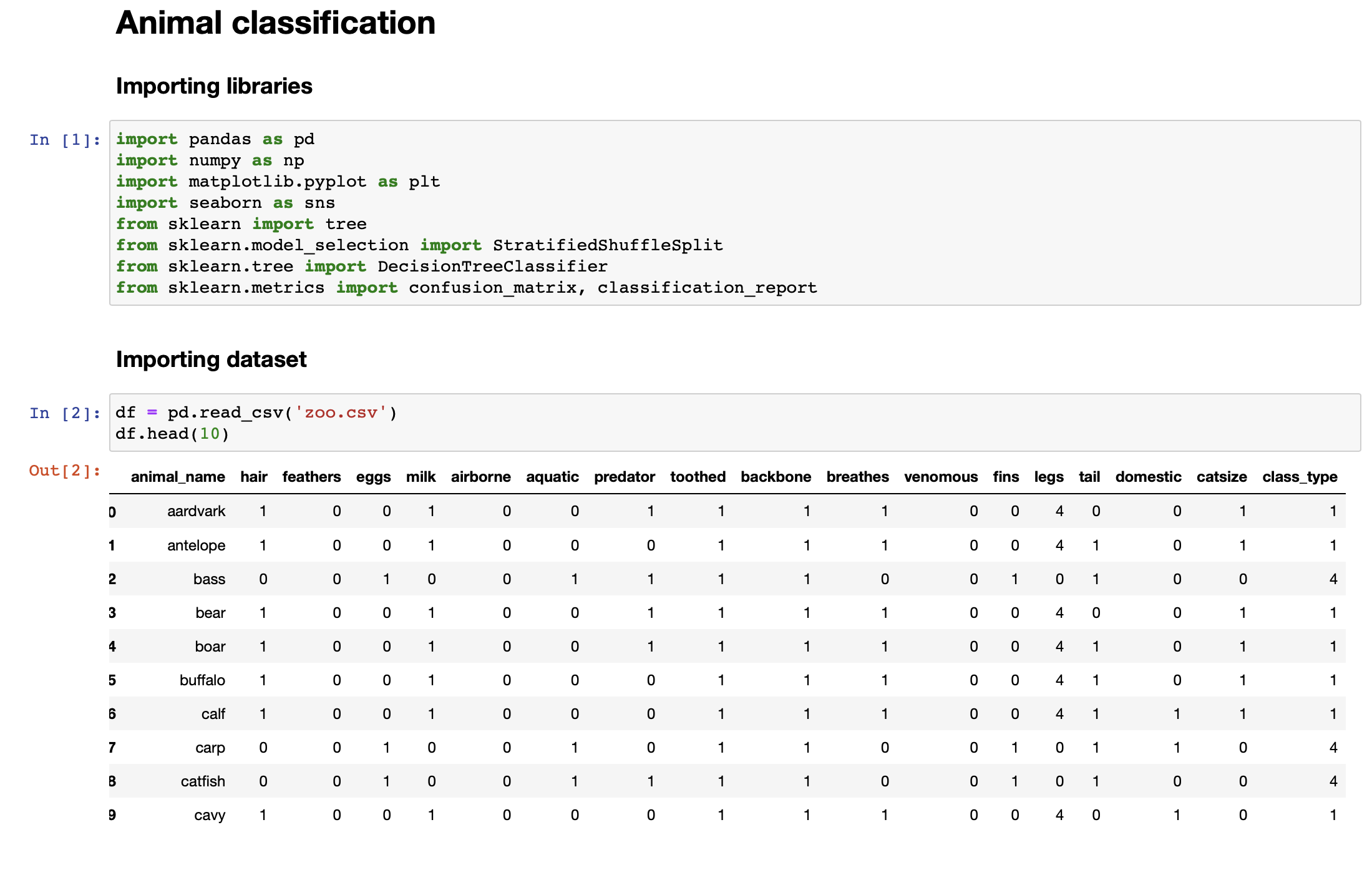
* **Feature Direction**

It will be a good chance if we try the experiment on another algorithm like SVM or to locate and recognize any animal on any given picture.

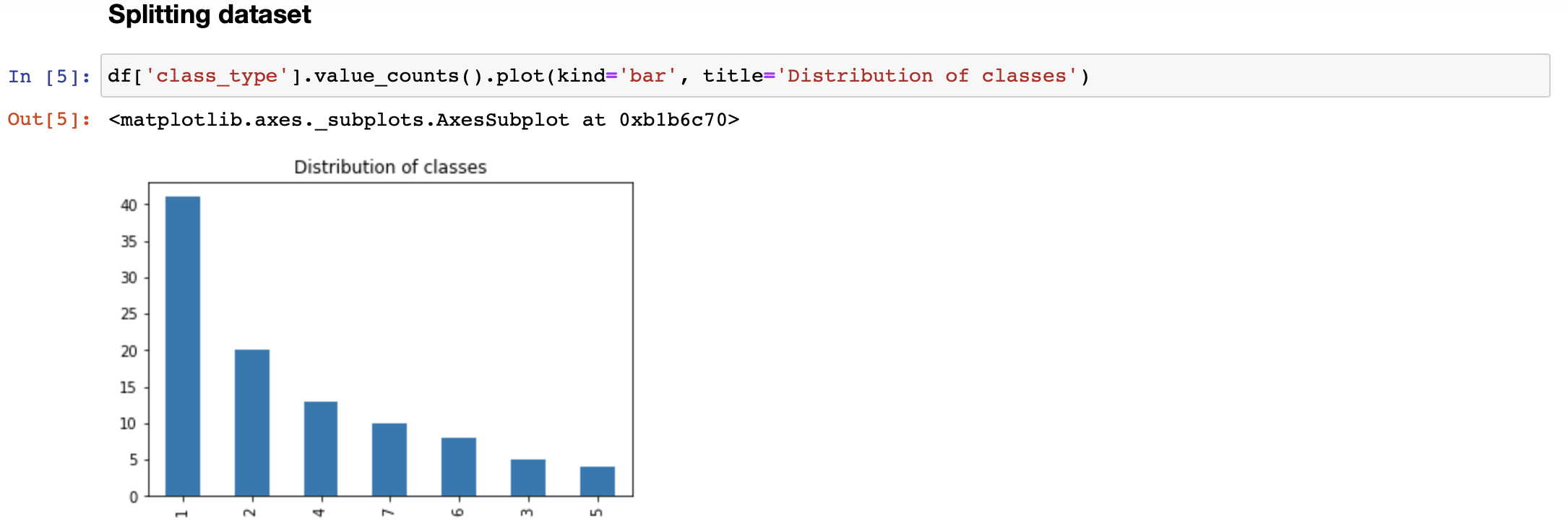
**Appendix**

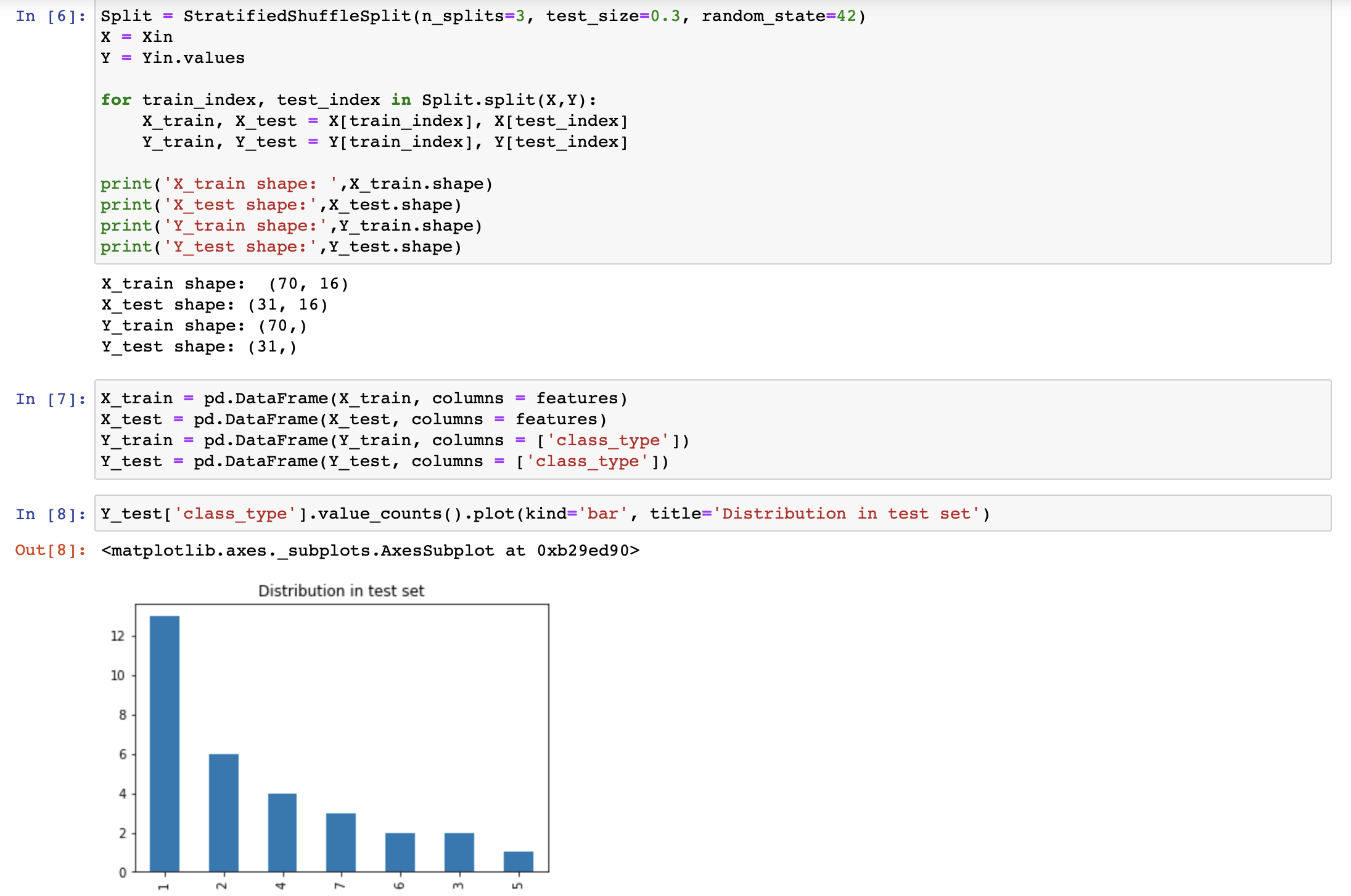
* Snapshots and Others

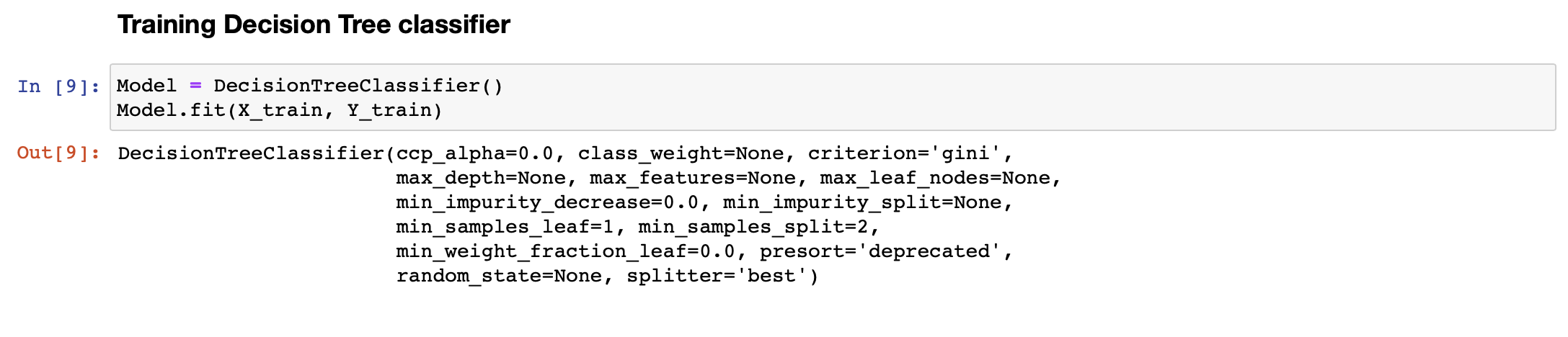
These snapshots represent the whole experiment:

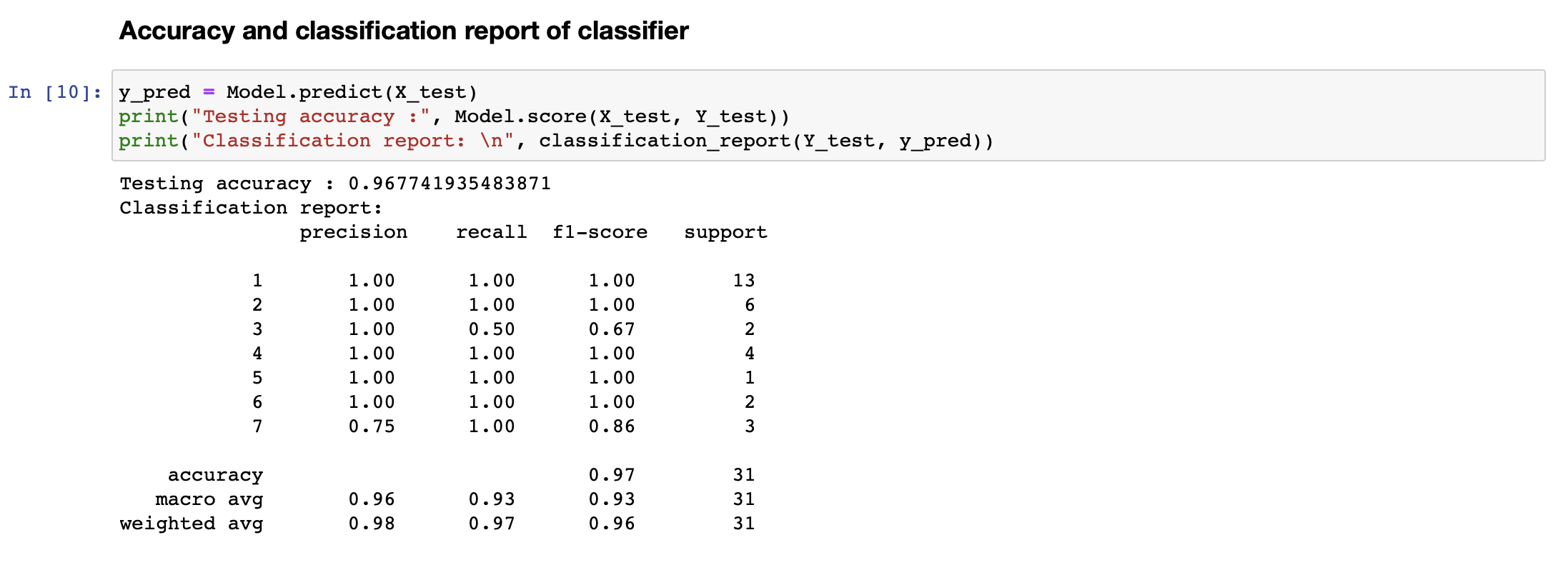


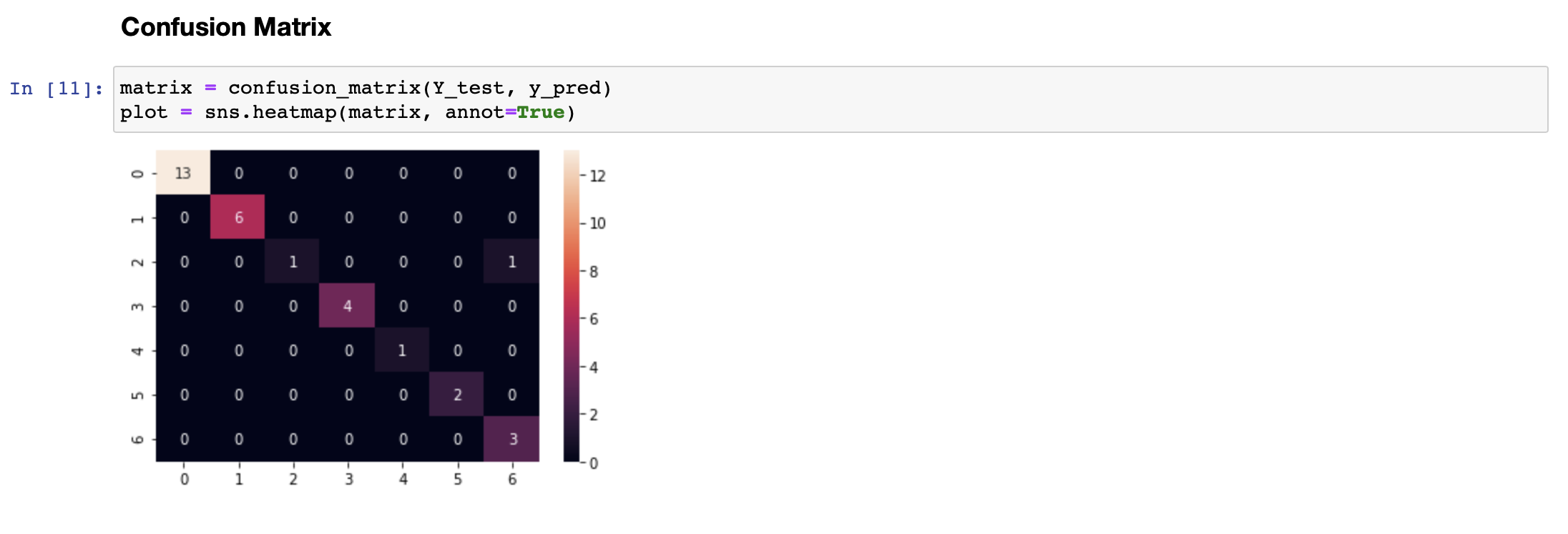














**References**

* <https://courses.lumenlearning.com/suny-biology2xmaster/chapter/features-used-to-classify-animals/>
* <https://www.kaggle.com/uciml/zoo-animal-classification#zoo.csv>
* <https://www.kaggle.com/carrie1/decision-tree-classification-using-zoo-animals>