**Problem**

The proposed problem is to create a model to classify animals into one of the following:

|  |
| --- |
| 1. Mammal |
| 1. Bird |
| 1. Reptile |
| 1. Fish |
| 1. Amphibian |
| 1. Bug |
| 1. Invertebrate |

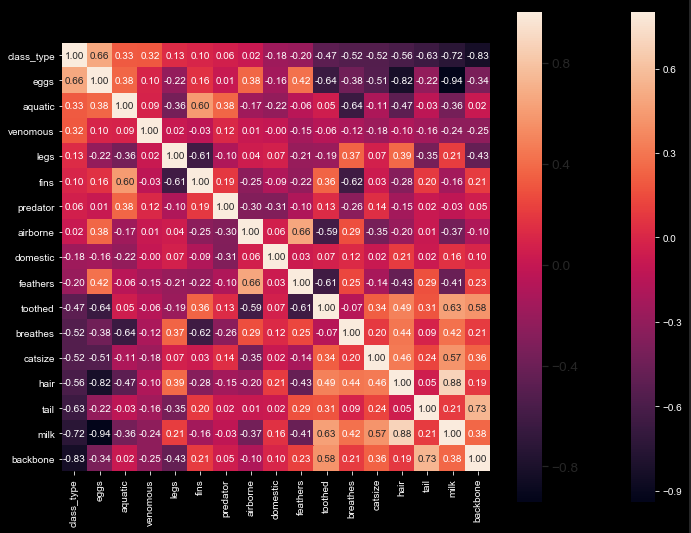
**Data**

The dataset contains 101 observations (animals) and no missing data. Except for animal\_name, all of the fields are binary, indicating an animal does or does not have a feature.

|  |  |
| --- | --- |
| animal\_name | Object |
| hair | Boolean |
| feathers | Boolean |
| eggs | Boolean |
| milk | Boolean |
| airborne | Boolean |
| aquatic | Boolean |
| predator | Boolean |
| toothed | Boolean |
| backbone | Boolean |
| breathes | Boolean |
| venomous | Boolean |
| fins | Boolean |
| legs | Boolean |
| tail | Boolean |
| domestic | Boolean |
| catsize | Boolean |
| class\_type | Boolean |

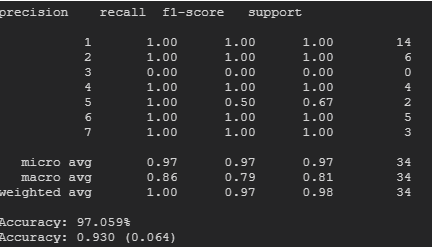
**Analysis**

I created two classification models for the dataset to find the best fit. Both models were cross validated with 10-fold validation. The dataset was split 33% for testing with the remainder used for training.



Model 1: Decision Tree

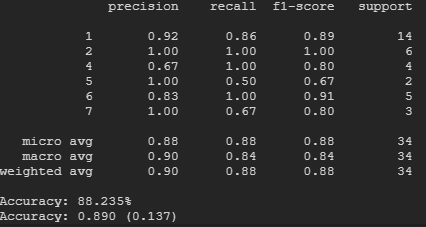
Due to assumptions about the data being minimal, all Boolean fields were entered as dependent variables into the model. This yielded a very high overall score. The only exception was for category 5 where the recall (ability of the classifier to find all the positive samples) appears low at 0.5. I believe this can be set aside as there are only two animals in the class and only four in the entire dataset.



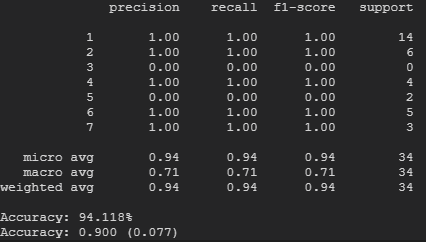
Model 2: K-Nearest Neighbor (KNN)

To ready the data for the KNN model, I removed all fields with less than an absolution value of 0.5 in correlation to the class\_type. I also removed the field hair due its it collinearity to milk (0.88). To prove these data alterations were effective, see the before and after classification scores below.

Before



After



**Results**

The decision tree model had the best score at 93% accuracy and should be chosen as the best fit for the data.