**Random Forest Classifier On Iris Dataset**

Michael Stewart

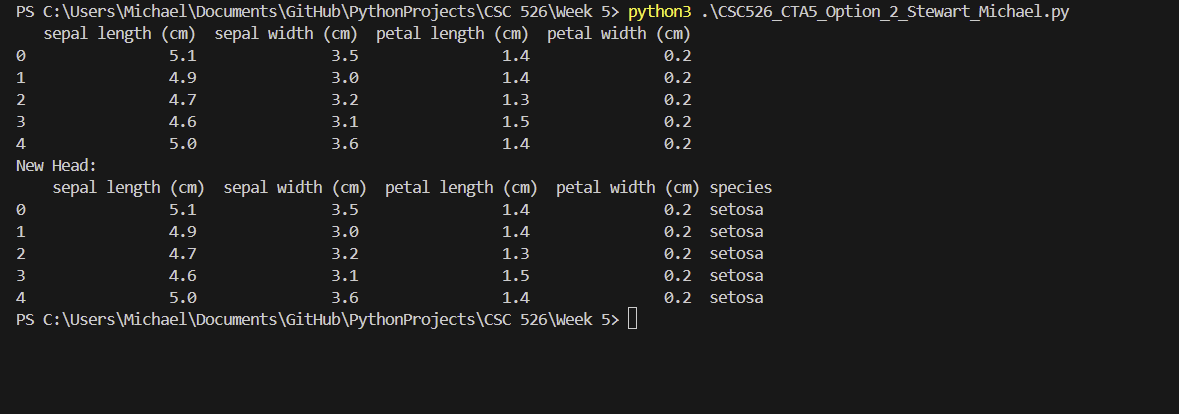
Colorado State University Global

CSC580-1: Applying Machine Learning and Neural Networks - Capstone

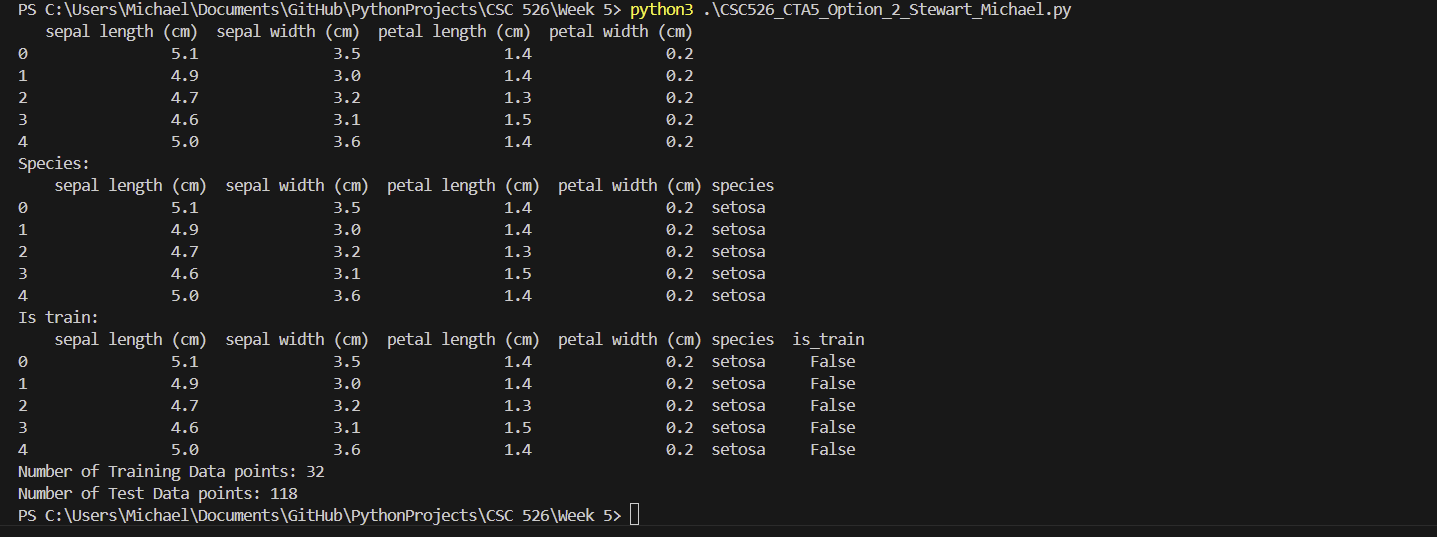
Professor Isaac Gang

October 20th, 2024

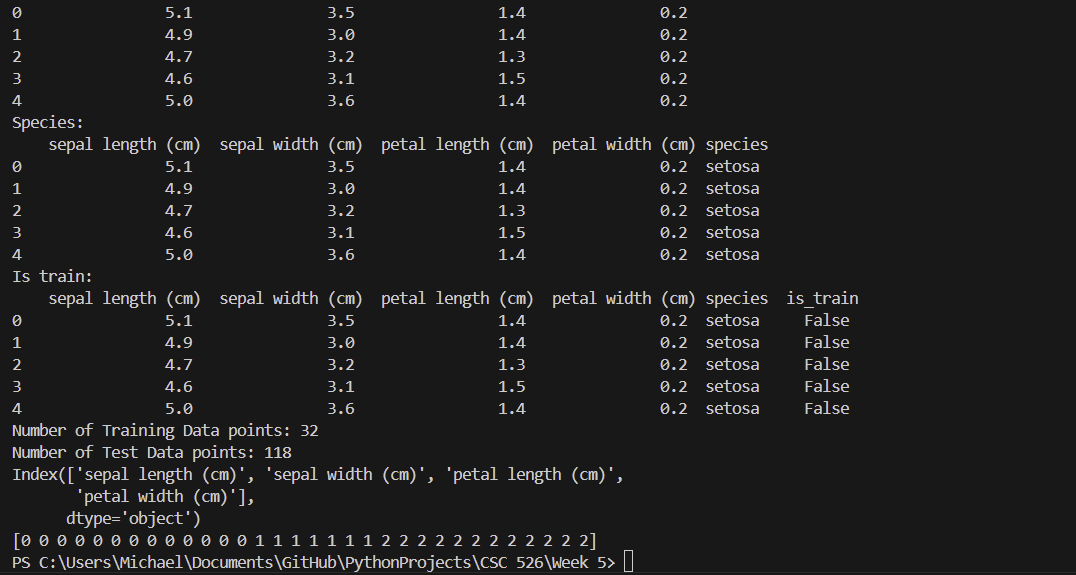
**Random Forest Classifier On Iris Dataset**

****

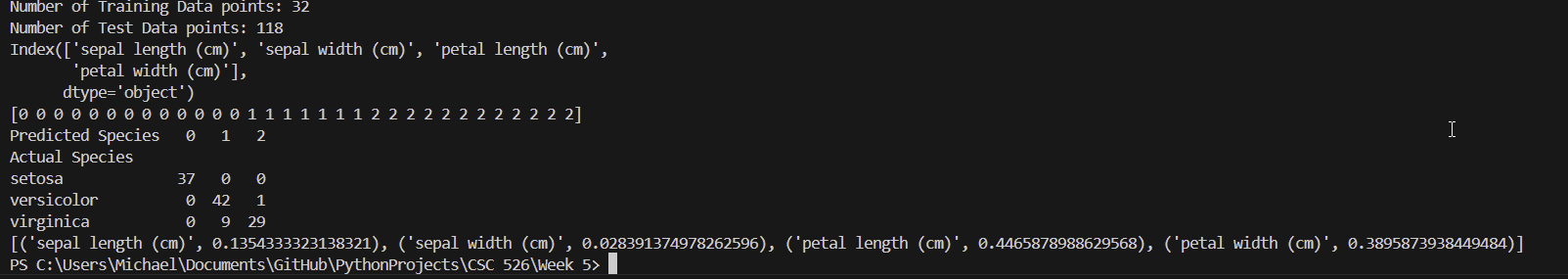
Here is the first screenshot after the first couple of steps. I imported the iris dataset and added the species column with the target. These are only the top 5 entries of the data frame using the df.head() function.

****

I then set aside about 25% of the data as training data and 75% as test data. This is similar to common machine learning and artificial intelligence techniques.

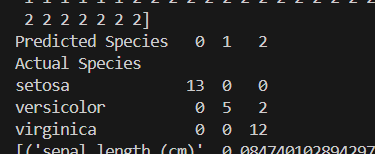
****

Pictured above are the number of training and test data points, features of the different irises, and the training data targets.

****

**[('sepal length (cm)', 0.1354333323138321), ('sepal width (cm)', 0.028391374978262596), ('petal length (cm)', 0.4465878988629568), ('petal width (cm)', 0.3895873938449484)]**

Finally, these are the results of our random forest classifier. You can see that the graph shows what each iris species was predicted and if it was correct. All of the data points should be on the diagonal, but there are 9 virginica flowers that were predicted to be versicolor and one versicolor was predicted to be virginica. An interesting experiment I tried was to use 75% of the data for training and 25% for testing. The accuracy was higher with only two missed measurements amounting to 94% accuracy compared to 93.5% accuracy of the previous one.



**References**

Ganegedara, T. (2022). *TensorFlow in action*. Simon and Schuster.