Applied Data Science – Assignment

Goal: Using a decision tree model in SKLearn library

This assignment has two parts, and both are pretty much like what we have learned in SKLearn presentations for recognizing if the fruit is an apple or an orange or if a flower is Iris Setosa, Iris Virginica or Iris Versicolor.

Part A:

We want to play outdoor tennis. Some weather conditions, such as low temperature, rainy or very windy ones prohibit us from playing. These are just a few examples, but you can find more features (columns). Please create a dataset, including at least three features and twenty samples (rows), and the labels could be cancelling or playing (like Yes/No or True/False).

After creating your model and training your classifier, using the above 20 rows, test your model over three new (unseen) samples, and see if it works properly or not.

Part B:

Please find/create a CSV file that has at least three columns (features) and 30 rows that can decide if a person has COVID-19 or not. Please use 30% of the dataset as your testing data and show the achieved accuracy. What we have learned in 02.pptx, and 04.pptx are very helpful and important to split our dataset (especially 04.pptx).