Course: Introduction to Data Science (DS2006) - Laboratory 14

Task 1:
The first line of the Iris dataset includes the names describing the different attributes being measured. The data measures the sepal and petal widths and lengths, as well as the classification of the flower.

Task 2:
From line 2 until the end illustrates the data samples collected for each attribute of different Iris flowers. A multitude of each species of Iris flowers are measured, and each species has multiple samples in order to collect a comprehensive dataset.

Task 3: knn.py implemented

Task 4:

	sepal.length	sepal.width	petal.length	petal.width	class
0	5.1	3.5	1.4	0.2	Setosa
1	4.9	3.0	1.4	0.2	Setosa
2	4.7	3.2	1.3	0.2	Setosa
3	4.6	3.1	1.5	0.2	Setosa
4	5.0	3.6	1.4	0.2	Setosa

Task 5: features and classes dataframes created

Task 6:

```
Features:
   sepal.length  sepal.width  petal.length  petal.width
0           5.1          3.5           1.4          0.2
1           4.9          3.0           1.4          0.2
2           4.7          3.2           1.3          0.2
3           4.6          3.1           1.5          0.2
4           5.0          3.6           1.4          0.2
```

```
Classes:
0    Setosa
1    Setosa
2    Setosa
3    Setosa
4    Setosa
Name: class, dtype: object
```

Task 7: train and test sets created for features and classes

Task 8: KNeighborsClassifier object created and trained

Task 9: predictions for the test gotten

Task 10:

```
Predictions: ['Versicolor' 'Virginica' 'Setosa' 'Versicolor' 'Setosa' 'Versicolor'
 'Virginica' 'Versicolor' 'Setosa' 'Versicolor' 'Versicolor' 'Virginica'
 'Versicolor' 'Setosa' 'Setosa' 'Virginica' 'Versicolor' 'Setosa' 'Setosa'
 'Setosa' 'Virginica' 'Virginica' 'Virginica' 'Setosa' 'Versicolor'
 'Setosa' 'Versicolor' 'Versicolor' 'Versicolor' 'Virginica']

Accuracy: 0.9666666666666667
```

Task 11:

```
              precision    recall  f1-score   support

      Setosa       1.00      1.00      1.00        10
  Versicolor       1.00      0.92      0.96        13
   Virginica       0.88      1.00      0.93         7

    accuracy                           0.97        30
   macro avg       0.96      0.97      0.96        30
weighted avg       0.97      0.97      0.97        30
```

Task 12: stratified split on the test set implemented

Task 13: strat-knn.py completed

Task 14: decision-tree.py submitted