Hands-on : Classification

In this hands on activity you will learn how to develop an end to end ML based classification. project. We will work with the iris flower dataset. The dataset contains features of a particular flower and the task is to classify it into one of the flower species namely Iris setosa, Iris virginica and Iris versicolor

The dataset filename is `iris.data.csv`

Attribute Information:

1. sepal length in cm
2. sepal width in cm
3. petal length in cm
4. petal width in cm
5. class: -- Iris Setosa -- Iris Versicolour -- Iris Virginica

1. Load the necessary libraries

In [2]:

```
# Fill in with your code
```

2. Load the dataset

In [3]:

```
# Fill in with your code
```

3. Summarize the dataset

In [4]:

```
# Fill in with your code
```

```
(150, 5)
```

The dataset contains 150 instances and 5 attributes

In [5]:

```
# Fill in with your code
```

```
    sepal-length  sepal-width  petal-length  petal-width        clas
s
0            5.1          3.5           1.4          0.2  Iris-setos
a
1            4.9          3.0           1.4          0.2  Iris-setos
a
2            4.7          3.2           1.3          0.2  Iris-setos
a
3            4.6          3.1           1.5          0.2  Iris-setos
a
4            5.0          3.6           1.4          0.2  Iris-setos
a
5            5.4          3.9           1.7          0.4  Iris-setos
a
6            4.6          3.4           1.4          0.3  Iris-setos
a
7            5.0          3.4           1.5          0.2  Iris-setos
a
8            4.4          2.9           1.4          0.2  Iris-setos
a
9            4.9          3.1           1.5          0.1  Iris-setos
a
10           5.4          3.7           1.5          0.2  Iris-setos
a
11           4.8          3.4           1.6          0.2  Iris-setos
a
12           4.8          3.0           1.4          0.1  Iris-setos
a
13           4.3          3.0           1.1          0.1  Iris-setos
a
14           5.8          4.0           1.2          0.2  Iris-setos
a
15           5.7          4.4           1.5          0.4  Iris-setos
a
16           5.4          3.9           1.3          0.4  Iris-setos
a
17           5.1          3.5           1.4          0.3  Iris-setos
a
18           5.7          3.8           1.7          0.3  Iris-setos
a
19           5.1          3.8           1.5          0.3  Iris-setos
a
```

The first 20 instances in the data

In [6]:

```
# Fill in with your code
```

```
       sepal-length  sepal-width  petal-length  petal-width
count    150.000000   150.000000    150.000000   150.000000
mean       5.843333     3.054000      3.758667     1.198667
std        0.828066     0.433594      1.764420     0.763161
min        4.300000     2.000000      1.000000     0.100000
25%        5.100000     2.800000      1.600000     0.300000
50%        5.800000     3.000000      4.350000     1.300000
75%        6.400000     3.300000      5.100000     1.800000
max        7.900000     4.400000      6.900000     2.500000
```

Statistical descriptions of the data

In [7]:

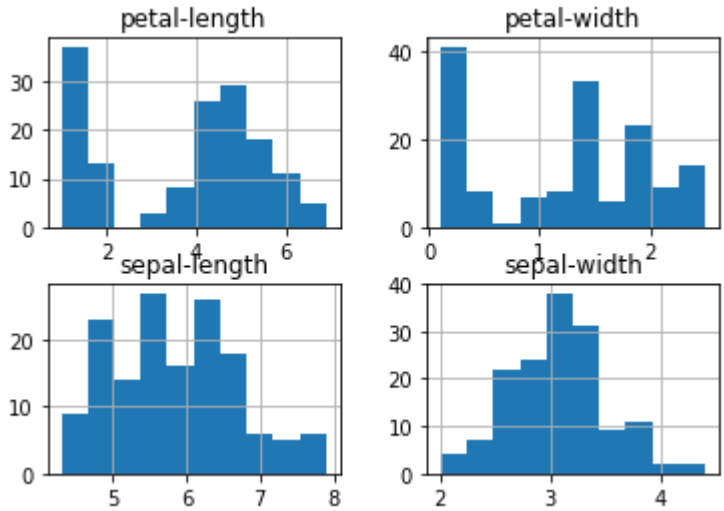
```
# Fill in with your code
```

```
class
Iris-setosa        50
Iris-versicolor    50
Iris-virginica     50
dtype: int64
```

Distribution of the class/label in the data

In [8]:

```
# Fill in with your code
```



4. Create a train test split

In [ ]:

```
# Fill in with your code
```

5. Cross Validation

Try using any classifier you like e.g. Logistic Regression then print the average precision and also the precision, recall, and F1 for each label

In [11]:

```
# Fill in with your code
```

```
0.950990990991
                 precision    recall  f1-score   support

    Iris-setosa       1.00      1.00      1.00        43
Iris-versicolor       0.97      0.86      0.91        42
 Iris-virginica       0.88      0.98      0.93        45

    avg / total       0.95      0.95      0.95       130
```

6. Evaluate performance on the test set

Now we have performed the cross validation, we can build the model on all training data and then test it on unseen data. Print the confusion matrix and also the classification report

In [12]:

```
# Fill in with your code
```

```
[[7 0 0]
 [0 3 5]
 [0 0 5]]
                 precision    recall  f1-score   support

    Iris-setosa       1.00      1.00      1.00         7
Iris-versicolor       1.00      0.38      0.55         8
 Iris-virginica       0.50      1.00      0.67         5

    avg / total       0.88      0.75      0.73        20
```

What if you want to compare several models? How would you change your code

In [ ]:

```

```