

In [1]:

```
1 import sys
2 print('Python={}'.format(sys.version))
3 import pandas
4 print('Pandas={}'.format(pandas.__version__))
5 import matplotlib
6 print('Matplotlib={}'.format(matplotlib.__version__))
7 import numpy
8 print('Numpy={}'.format(numpy.__version__))
9 import scipy
10 print('Scipy={}'.format(scipy.__version__))
11 import sklearn
12 print('Sklearn={}'.format(sklearn.__version__))
```

Python=3.9.7 (default, Sep 16 2021, 16:59:28) [MSC v.1916 64 bit (AMD64)]
Pandas=1.3.4
Matplotlib=3.4.3
Numpy=1.20.3
Scipy=1.7.1
Sklearn=0.24.2

In [2]:

```
1 import pandas as pd
2 from pandas.plotting import scatter_matrix
3 from matplotlib import pyplot
4 from sklearn.model_selection import train_test_split
5 from sklearn.model_selection import cross_val_score
6 from sklearn.model_selection import StratifiedKFold
7 from sklearn.metrics import classification_report
8 from sklearn.metrics import confusion_matrix
9 from sklearn.metrics import accuracy_score
10 from sklearn.linear_model import LogisticRegression
11 from sklearn.tree import DecisionTreeClassifier
12 from sklearn.neighbors import KNeighborsClassifier
13 from sklearn.discriminant_analysis import LinearDiscriminantAnalysis
14 from sklearn.naive_bayes import GaussianNB
15 from sklearn.svm import SVC
16 from sklearn import model_selection
17 from sklearn.ensemble import VotingClassifier
```

In [3]:

```
1 dataset=pd.read_csv('iris.csv')
```

In [4]:

```
1 print(dataset.shape)
```

(150, 5)

In [5]:

```
1 print(dataset.head(20))
```

	sepal.length	sepal.width	petal.length	petal.width	variety
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
5	5.4	3.9	1.7	0.4	Setosa
6	4.6	3.4	1.4	0.3	Setosa
7	5.0	3.4	1.5	0.2	Setosa
8	4.4	2.9	1.4	0.2	Setosa
9	4.9	3.1	1.5	0.1	Setosa
10	5.4	3.7	1.5	0.2	Setosa
11	4.8	3.4	1.6	0.2	Setosa
12	4.8	3.0	1.4	0.1	Setosa
13	4.3	3.0	1.1	0.1	Setosa
14	5.8	4.0	1.2	0.2	Setosa
15	5.7	4.4	1.5	0.4	Setosa
16	5.4	3.9	1.3	0.4	Setosa
17	5.1	3.5	1.4	0.3	Setosa
18	5.7	3.8	1.7	0.3	Setosa
19	5.1	3.8	1.5	0.3	Setosa

In [6]:

```
1 print(dataset.describe())
```

	sepal.length	sepal.width	petal.length	petal.width
count	150.000000	150.000000	150.000000	150.000000
mean	5.843333	3.057333	3.758000	1.199333
std	0.828066	0.435866	1.765298	0.762238
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

In [7]:

```
1 print(dataset.groupby('variety').size())
```

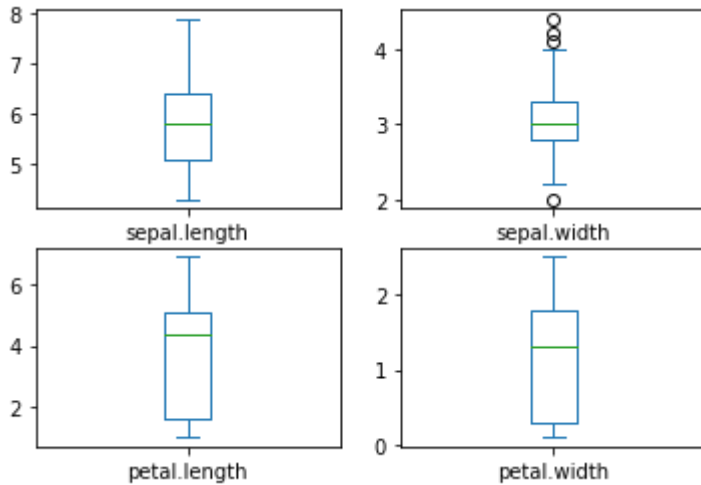
```
variety
Setosa      50
Versicolor  50
Virginica   50
dtype: int64
```

In [8]:

```
1 dataset.plot(kind='box',subplots=True,layout=(2,2),sharex=False,sharey=False)
```

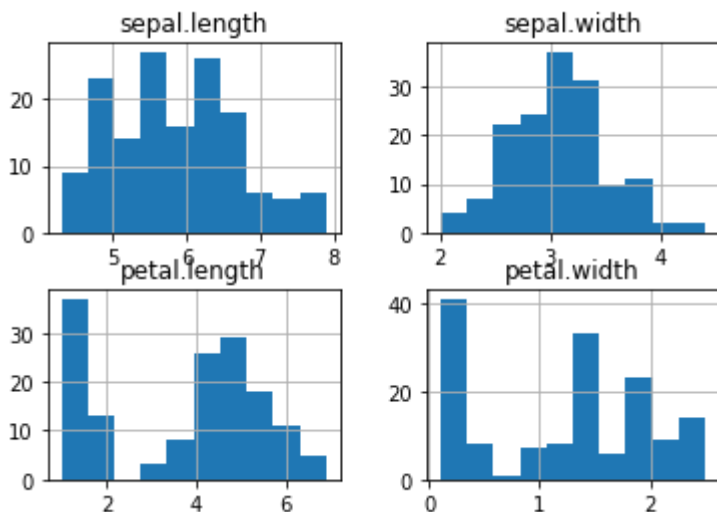
Out[8]:

```
sepal.length    AxesSubplot(0.125,0.536818;0.352273x0.343182)  
sepal.width     AxesSubplot(0.547727,0.536818;0.352273x0.343182)  
petal.length    AxesSubplot(0.125,0.125;0.352273x0.343182)  
petal.width     AxesSubplot(0.547727,0.125;0.352273x0.343182)  
dtype: object
```



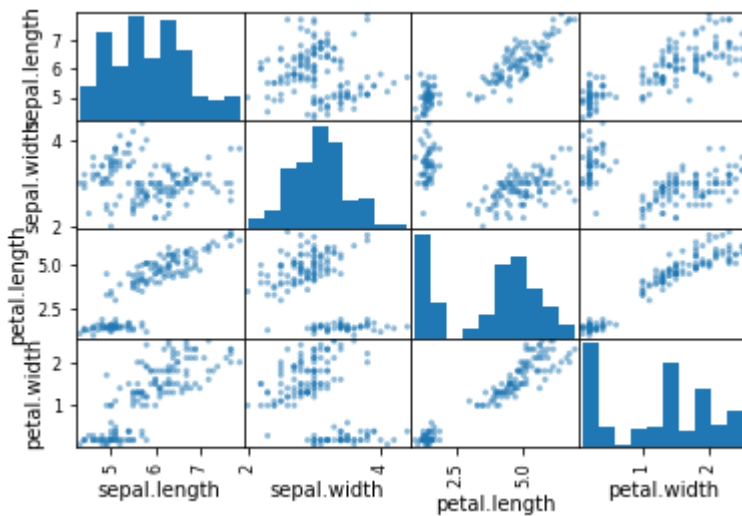
In [9]:

```
1 dataset.hist()  
2 pyplot.show()
```



In [10]:

```
1 scatter_matrix(dataset)
2 pyplot.show()
```



In [11]:

```
1 array=dataset.values
2 X=array[:,0:4]
3 Y=array[:,4]
4 X_train,X_validation,Y_train,Y_validation=train_test_split(X,Y,test_size=0.2,random.
```

In [12]:

```
1 models=[]
2 models.append(('LR',LogisticRegression(solver='liblinear',multi_class='ovr')))
3 models.append(('LDA',LinearDiscriminantAnalysis()))
4 models.append(('KNN',KNeighborsClassifier()))
5 models.append(('NB',GaussianNB()))
6 models.append(('SVM',SVC(gamma='auto')))
```

In [13]:

```
1 result=[]
2 name=[]
3 for n,m in models:
4     kfold=StratifiedKFold(n_splits=10,shuffle=True,random_state=1)
5     cv_results=cross_val_score(m,X_train,Y_train,cv=kfold,scoring='accuracy')
6     result.append(cv_results)
7     name.append(n)
8     print('%s:%f (%f)' % (n,cv_results.mean(),cv_results.std()))
```

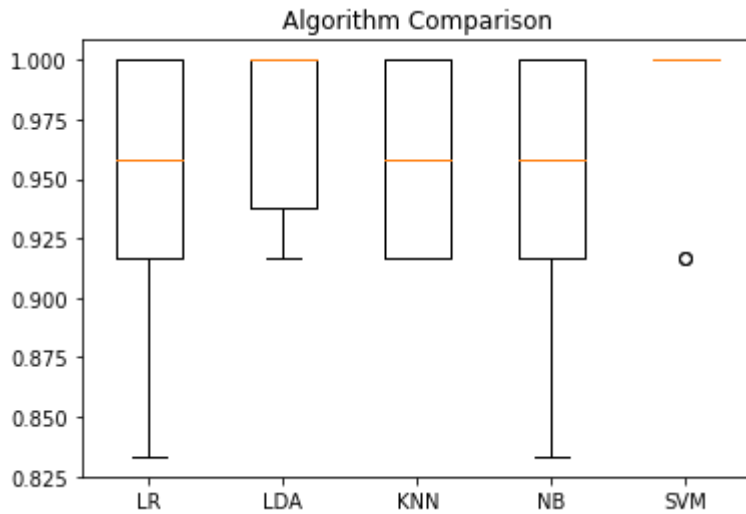
```
LR:0.941667 (0.065085)
LDA:0.975000 (0.038188)
KNN:0.958333 (0.041667)
NB:0.950000 (0.055277)
SVM:0.983333 (0.033333)
```

In [14]:

```

1 pyplot.boxplot(result,labels=name)
2 pyplot.title('Algorithm Comparison')
3 pyplot.show()

```



In [15]:

```

1 model=SVC(gamma='auto')
2 model.fit(X_train,Y_train)
3 pred=model.predict(X_validation)

```

In [16]:

```

1 print(accuracy_score(Y_validation,pred))
2 print(confusion_matrix(Y_validation,pred))
3 print(classification_report(Y_validation,pred))

```

0.9666666666666667

```

[[11  0  0]
 [ 0 12  1]
 [ 0  0  6]]

```

	precision	recall	f1-score	support
Setosa	1.00	1.00	1.00	11
Versicolor	1.00	0.92	0.96	13
Virginica	0.86	1.00	0.92	6
accuracy			0.97	30
macro avg	0.95	0.97	0.96	30
weighted avg	0.97	0.97	0.97	30