In [1]:
```
#import libraries
import numpy as np
import pandas as pd
import seaborn as sns
from sklearn.preprocessing import LabelEncoder
from sklearn.model_selection import train_test_split
```

In [2]:
```
#to access file
df=pd.read_csv("C:\\Users\\Administrator\\Documents\\ml\\archive (41)\\IRIS.csv")
df
```

Out[2]:

	sepal_length	sepal_width	petal_length	petal_width	species
0	5.1	3.5	1.4	0.2	Iris-setosa
1	4.9	3.0	1.4	0.2	Iris-setosa
2	4.7	3.2	1.3	0.2	Iris-setosa
3	4.6	3.1	1.5	0.2	Iris-setosa
4	5.0	3.6	1.4	0.2	Iris-setosa
...	...	...	...	...	...
145	6.7	3.0	5.2	2.3	Iris-virginica
146	6.3	2.5	5.0	1.9	Iris-virginica
147	6.5	3.0	5.2	2.0	Iris-virginica
148	6.2	3.4	5.4	2.3	Iris-virginica
149	5.9	3.0	5.1	1.8	Iris-virginica

150 rows × 5 columns

In [3]:
```
#species has three class so we assign label to them
le=LabelEncoder()
speciess=le.fit_transform(df['species'])
```

In [4]:
```
#drop species column from dataset
df2=df.drop('species',axis=1)
df2
```

Out[4]:

	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
...	...	...	...	...
145	6.7	3.0	5.2	2.3
146	6.3	2.5	5.0	1.9
147	6.5	3.0	5.2	2.0
148	6.2	3.4	5.4	2.3
149	5.9	3.0	5.1	1.8

150 rows × 4 columns

In [5]:
```
#add new column speciess in which we have done label encoding
df2['speciess']=speciess
df2
```

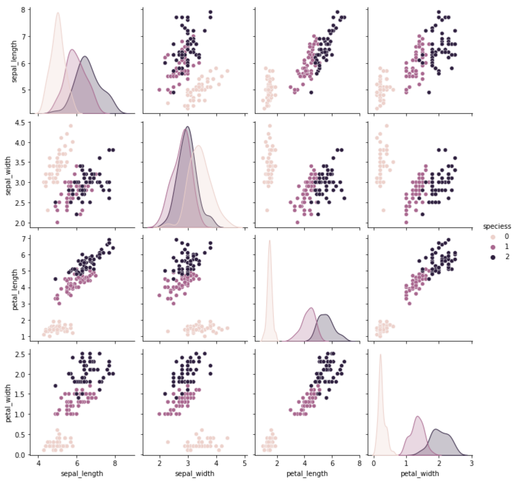
Out[5]:

	sepal_length	sepal_width	petal_length	petal_width	speciess
0	5.1	3.5	1.4	0.2	0
1	4.9	3.0	1.4	0.2	0
2	4.7	3.2	1.3	0.2	0
3	4.6	3.1	1.5	0.2	0
4	5.0	3.6	1.4	0.2	0
...	...	...	...	...	...
145	6.7	3.0	5.2	2.3	2
146	6.3	2.5	5.0	1.9	2
147	6.5	3.0	5.2	2.0	2
148	6.2	3.4	5.4	2.3	2
149	5.9	3.0	5.1	1.8	2

150 rows × 5 columns

In [6]:
```
#data visualization using seaborn
sns.pairplot(df2,hue='speciess')
```

Out[6]:
```
<seaborn.axisgrid.PairGrid at 0x22de76bd190>
```



In [7]:
```
X=df2.drop('speciess',axis=1)
y=df2['speciess']
```

In [8]:
```
#train_test_split the dataset
X_train,X_test,y_train,y_test=train_test_split(X,y,test_size=0.2,random_state=51)
```

In [9]:
```
#import linear regression library
from sklearn.linear_model import LinearRegression
lr=LinearRegression()
lr.fit(X_train,y_train)
lr.score(X_test,y_test)
```

Out[9]:
```
0.9132273418532224
```

In [10]:
```
#import RandomForest classifier
from sklearn.ensemble import RandomForestClassifier
rf=RandomForestClassifier()
rf.fit(X_train,y_train)
rf.score(X_test,y_test)
```

Out[10]:
```
0.9333333333333333
```

In [11]:
```
#import knn classifier
from sklearn.neighbors import KNeighborsClassifier
rfl=KNeighborsClassifier()
rfl.fit(X_train,y_train)
rfl.score(X_test,y_test)
```

Out[11]:
```
0.9666666666666667
```

In [12]:
```
#predict X_test and compare it to y_test
pred=rfl.predict(X_test)
pred
```

Out[12]:
```
array([1, 2, 1, 0, 1, 2, 2, 2, 0, 2, 2, 0, 1, 2, 2, 1, 1, 2, 2, 0, 2, 2,
       0, 0, 1, 2, 0, 0, 0, 2])
```

In [13]:
```
y_test
```

Out[13]:
```
78     1
121    2
99     1
36     0
92     1
107    2
131    2
144    2
10     0
126    2
127    2
4      0
73     1
123    2
136    2
70     1
119    2
100    2
117    2
43     0
145    2
111    2
31     0
45     0
58     1
135    2
21     0
14     0
15     0
101    2
Name: speciess, dtype: int32
```

In [ ]:
```
```