In [5]:
```
import numpy as pd
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.neighbors import KNeighborsClassifier
from sklearn.metrics import accuracy_score
```

In [10]:
```
df = pd.read_csv(r"Iris.csv")
```

In [11]:
```
print (df)
```

```
      Id  SepalLengthCm  SepalWidthCm  PetalLengthCm  PetalWidthCm  \
0      1            5.1           3.5            1.4           0.2
1      2            4.9           3.0            1.4           0.2
2      3            4.7           3.2            1.3           0.2
3      4            4.6           3.1            1.5           0.2
4      5            5.0           3.6            1.4           0.2
..   ...            ...           ...            ...           ...
145  146            6.7           3.0            5.2           2.3
146  147            6.3           2.5            5.0           1.9
147  148            6.5           3.0            5.2           2.0
148  149            6.2           3.4            5.4           2.3
149  150            5.9           3.0            5.1           1.8

            Species
0       Iris-setosa
1       Iris-setosa
2       Iris-setosa
3       Iris-setosa
4       Iris-setosa
..              ...
145  Iris-virginica
146  Iris-virginica
147  Iris-virginica
148  Iris-virginica
149  Iris-virginica

[150 rows x 6 columns]
```

In [12]:
```
df.head
```

Out[12]:
```
<bound method NDFrame.head of       Id  SepalLengthCm  SepalWidthCm  PetalLen
gthCm  PetalWidthCm  \
0      1            5.1           3.5            1.4           0.2
1      2            4.9           3.0            1.4           0.2
2      3            4.7           3.2            1.3           0.2
3      4            4.6           3.1            1.5           0.2
4      5            5.0           3.6            1.4           0.2
..   ...            ...           ...            ...           ...
145  146            6.7           3.0            5.2           2.3
146  147            6.3           2.5            5.0           1.9
147  148            6.5           3.0            5.2           2.0
148  149            6.2           3.4            5.4           2.3
149  150            5.9           3.0            5.1           1.8

            Species
0       Iris-setosa
1       Iris-setosa
2       Iris-setosa
3       Iris-setosa
4       Iris-setosa
..              ...
145  Iris-virginica
146  Iris-virginica
147  Iris-virginica
148  Iris-virginica
149  Iris-virginica

[150 rows x 6 columns]>
```

In [13]:
```
df.tail
```

Out[13]:
```
<bound method NDFrame.tail of       Id  SepalLengthCm  SepalWidthCm  PetalLen
gthCm  PetalWidthCm  \
0      1            5.1           3.5            1.4           0.2
1      2            4.9           3.0            1.4           0.2
2      3            4.7           3.2            1.3           0.2
3      4            4.6           3.1            1.5           0.2
4      5            5.0           3.6            1.4           0.2
..   ...            ...           ...            ...           ...
145  146            6.7           3.0            5.2           2.3
146  147            6.3           2.5            5.0           1.9
147  148            6.5           3.0            5.2           2.0
148  149            6.2           3.4            5.4           2.3
149  150            5.9           3.0            5.1           1.8

            Species
0       Iris-setosa
1       Iris-setosa
2       Iris-setosa
3       Iris-setosa
4       Iris-setosa
..              ...
145  Iris-virginica
146  Iris-virginica
147  Iris-virginica
148  Iris-virginica
149  Iris-virginica

[150 rows x 6 columns]>
```

In [14]:
```
df['PetalWidthCm']
```

Out[14]:
```
0      0.2
1      0.2
2      0.2
3      0.2
4      0.2
      ...
145    2.3
146    1.9
147    2.0
148    2.3
149    1.8
Name: PetalWidthCm, Length: 150, dtype: float64
```

In [15]:
```
projected_columns = ['PetalWidthCm','Species']
df[projected_columns]
```

Out[15]:

	PetalWidthCm	Species
0	0.2	Iris-setosa
1	0.2	Iris-setosa
2	0.2	Iris-setosa
3	0.2	Iris-setosa
4	0.2	Iris-setosa
...	...	...
145	2.3	Iris-virginica
146	1.9	Iris-virginica
147	2.0	Iris-virginica
148	2.3	Iris-virginica
149	1.8	Iris-virginica

150 rows × 2 columns

In [16]:
```
df['Species'].value_counts()
```

Out[16]:
```
Iris-setosa        50
Iris-versicolor    50
Iris-virginica     50
Name: Species, dtype: int64
```

In [17]:
```
input_features = ['SepalLengthCm', 'SepalWidthCm', 'PetalLengthCm', 'PetalWidt
df_input = df[input_features]
df_input
```

Out[17]:

	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm
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 [18]:
```
output_features = ['Species']
df_output = df[output_features]
df_output
```

Out[18]:

	Species
0	Iris-setosa
1	Iris-setosa
2	Iris-setosa
3	Iris-setosa
4	Iris-setosa
...	...
145	Iris-virginica
146	Iris-virginica
147	Iris-virginica
148	Iris-virginica
149	Iris-virginica

150 rows × 1 columns

In [19]:
```
a, b, c = [2, 3, 1]
```

In [20]:
```
df_input_train, df_input_test, df_output_train, df_output_test = train_test_sp
```

In [21]:
```
df_output_test
```

Out[21]:

	Species
101	Iris-virginica
93	Iris-versicolor
133	Iris-virginica
144	Iris-virginica
89	Iris-versicolor
5	Iris-setosa
129	Iris-virginica
19	Iris-setosa
53	Iris-versicolor
43	Iris-setosa
114	Iris-virginica
11	Iris-setosa
71	Iris-versicolor
104	Iris-virginica
119	Iris-virginica
138	Iris-virginica
127	Iris-virginica
88	Iris-versicolor
115	Iris-virginica
18	Iris-setosa
7	Iris-setosa
79	Iris-versicolor
46	Iris-setosa
118	Iris-virginica
147	Iris-virginica
80	Iris-versicolor
111	Iris-virginica
56	Iris-versicolor
65	Iris-versicolor
69	Iris-versicolor

In [39]:
```
import sklearn.model_selection
import numpy as np
model = KNeighborsClassifier()
```

In [26]:
```
df_output_test
```

Out[26]:

	Species
101	Iris-virginica
93	Iris-versicolor
133	Iris-virginica
144	Iris-virginica
89	Iris-versicolor
5	Iris-setosa
129	Iris-virginica
19	Iris-setosa
53	Iris-versicolor
43	Iris-setosa
114	Iris-virginica
11	Iris-setosa
71	Iris-versicolor
104	Iris-virginica
119	Iris-virginica
138	Iris-virginica
127	Iris-virginica
88	Iris-versicolor
115	Iris-virginica
18	Iris-setosa
7	Iris-setosa
79	Iris-versicolor
46	Iris-setosa
118	Iris-virginica
147	Iris-virginica
80	Iris-versicolor
111	Iris-virginica
56	Iris-versicolor
65	Iris-versicolor
69	Iris-versicolor

In [29]:
```
data = {
    'Id': [23, 24, 12, 56, 87],
    'Gender': ['M', 'F', 'F', 'M', 'F']
}

df = pd.DataFrame(data)

gender_map = {
    'M':  0,
    'F': 1
}

df['Gender'] = df['Gender'].map(gender_map)

df.head()
```

Out[29]:

	Id	Gender
0	23	0
1	24	1
2	12	1
3	56	0
4	87	1

In [32]:
```
data = {
    'Age': [2, np.nan, 5, 1, 4],
    'Weight': [np.nan, 3, 6, np.nan, 1]
}

df = pd.DataFrame(data)

df.head()
```

Out[32]:

	Age	Weight
0	2.0	NaN
1	NaN	3.0
2	5.0	6.0
3	1.0	NaN
4	4.0	1.0

In [33]:
```
df.isnull().sum()
```

Out[33]:
```
Age       1
Weight    2
dtype: int64
```

In [ ]: