ML Lab-7 Assessment

In [1]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.neighbors import KNeighborsClassifier
```

In [2]:
```
fruits = pd.read_table('fruit_data_with_colors.txt')
```

In [3]:
```
fruits.head()
```

Out[3]:

	fruit_label	fruit_name	fruit_subtype	mass	width	height	color_score
0	1	apple	granny_smith	192	8.4	7.3	0.55
1	1	apple	granny_smith	180	8.0	6.8	0.59
2	1	apple	granny_smith	176	7.4	7.2	0.60
3	2	mandarin	mandarin	86	6.2	4.7	0.80
4	2	mandarin	mandarin	84	6.0	4.6	0.79

In [4]:
```
fruits.shape
```

Out[4]:
```
(59, 7)
```

In [5]:
```
predct = dict(zip(fruits.fruit_label.unique(), fruits.fruit_name.unique()))
predct
```

Out[5]:
```
{1: 'apple', 2: 'mandarin', 3: 'orange', 4: 'lemon'}
```

Data Preprocessing

In [6]:
```
fruits['fruit_name'].value_counts()
```

Out[6]:
```
apple       19
orange      19
lemon       16
mandarin     5
Name: fruit_name, dtype: int64
```

In [7]:
```
apple_data=fruits[fruits['fruit_name']=='apple']
orange_data=fruits[fruits['fruit_name']=='orange']
lemon_data=fruits[fruits['fruit_name']=='lemon']
mandarin_data=fruits[fruits['fruit_name']=='mandarin']
```

In [8]:
```
apple_data.head()
```

Out[8]:

	fruit_label	fruit_name	fruit_subtype	mass	width	height	color_score
0	1	apple	granny_smith	192	8.4	7.3	0.55
1	1	apple	granny_smith	180	8.0	6.8	0.59
2	1	apple	granny_smith	176	7.4	7.2	0.60
8	1	apple	braeburn	178	7.1	7.8	0.92
9	1	apple	braeburn	172	7.4	7.0	0.89

In [9]:
```
mandarin_data.head()
```

Out[9]:

	fruit_label	fruit_name	fruit_subtype	mass	width	height	color_score
3	2	mandarin	mandarin	86	6.2	4.7	0.80
4	2	mandarin	mandarin	84	6.0	4.6	0.79
5	2	mandarin	mandarin	80	5.8	4.3	0.77
6	2	mandarin	mandarin	80	5.9	4.3	0.81
7	2	mandarin	mandarin	76	5.8	4.0	0.81

In [10]:
```
orange_data.head()
```

Out[10]:

	fruit_label	fruit_name	fruit_subtype	mass	width	height	color_score
24	3	orange	spanish_jumbo	342	9.0	9.4	0.75
25	3	orange	spanish_jumbo	356	9.2	9.2	0.75
26	3	orange	spanish_jumbo	362	9.6	9.2	0.74
27	3	orange	selected_seconds	204	7.5	9.2	0.77
28	3	orange	selected_seconds	140	6.7	7.1	0.72

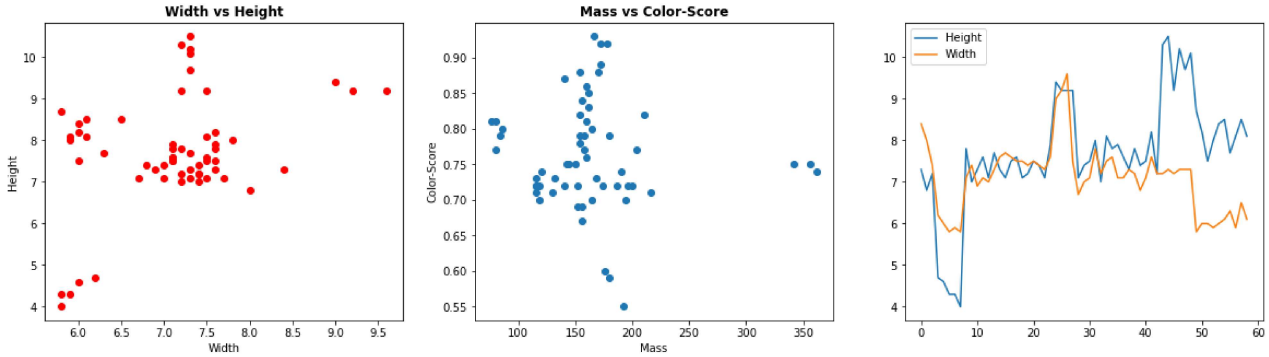
In [11]:
```
lemon_data.head()
```

Out[11]:

	fruit_label	fruit_name	fruit_subtype	mass	width	height	color_score
43	4	lemon	spanish_belsan	194	7.2	10.3	0.70
44	4	lemon	spanish_belsan	200	7.3	10.5	0.72
45	4	lemon	spanish_belsan	186	7.2	9.2	0.72
46	4	lemon	spanish_belsan	216	7.3	10.2	0.71
47	4	lemon	spanish_belsan	196	7.3	9.7	0.72

Data Visualisation

In [12]:
```
figure = plt.figure(figsize=(20, 5))
plt.subplot(1, 3, 1)
plt.scatter(fruits['width'],fruits['height'], color='r')
plt.title("Width vs Height", fontweight='bold')
plt.xlabel('Width')
plt.ylabel('Height')
plt.subplot(1, 3, 2)
plt.scatter(fruits['mass'],fruits['color_score'])
plt.title("Mass vs Color-Score", fontweight='bold')
plt.xlabel('Mass')
plt.ylabel('Color-Score')
plt.subplot(1, 3, 3)
plt.plot(fruits['height'],label='Height')
plt.plot(fruits['width'],label='Width')
plt.legend()
plt.show()
```



KNN Classifier

In [13]:
```
X=fruits[['mass','width','height']]
Y=fruits['fruit_label']
X_train,X_test,y_train,y_test=train_test_split(X,Y,random_state=0)
```

In [14]:
```
knn=KNeighborsClassifier()
```

In [15]:
```
knn.fit(X_train,y_train)
```

Out[15]:
```
KNeighborsClassifier()
```

In [16]:
```
knn.score(X_test,y_test)
```

Out[16]:
```
0.5333333333333333
```

Checking the Prediction

In [17]:
```
prediction1=knn.predict([[100,6.3,8]])
predct[prediction1[0]]
```

Out[17]: 'lemon'

In [18]:
```
prediction2=knn.predict([[300,7,10]])
predct[prediction2[0]]
```

Out[18]: 'orange'