Verzeo Internship - Minor Project - ML June Batch

Problem Statement :

To perform classification analysis on Iris dataset. Perform any two classification algorithms and compare the accuracy.

In [1]:

```
import pandas as pd
```

In [2]:

```
df = pd.read_csv("Iris.csv")
```

In [3]:

```
df
```

Out[3]:

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

150 rows × 6 columns

In [4]:

```
x = df.iloc[:, 1:5].values
y = df.iloc[:, 5].values
```

In [5]:

```
from sklearn.model_selection import train_test_split
```

In [6]:

```
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size = 0.3, random_st
```

In [7]:

```
x_test.shape
```

Out[7]:

```
(45, 4)
```

In [8]:

```
y_test.shape
```

Out[8]:

```
(45,)
```

Classification 1: Decision Tree Classification

In [9]:

```
from sklearn.tree import DecisionTreeClassifier
```

In [10]:

```
dt_classifier = DecisionTreeClassifier()
```

In [11]:

```
dt_classifier.fit(x_train, y_train)
```

Out[11]:

```
DecisionTreeClassifier()
```

In [12]:

```
y_pred = dt_classifier.predict(x_test)
y_pred
```

Out[12]:

```
array(['Iris-setosa', 'Iris-versicolor', 'Iris-versicolor', 'Iris-seto
sa',
       'Iris-virginica', 'Iris-versicolor', 'Iris-virginica',
       'Iris-setosa', 'Iris-setosa', 'Iris-virginica', 'Iris-versicolo
r',
       'Iris-setosa', 'Iris-virginica', 'Iris-versicolor',
       'Iris-versicolor', 'Iris-setosa', 'Iris-versicolor',
       'Iris-versicolor', 'Iris-setosa', 'Iris-setosa', 'Iris-versicol
or',
       'Iris-versicolor', 'Iris-virginica', 'Iris-setosa',
       'Iris-virginica', 'Iris-versicolor', 'Iris-setosa', 'Iris-setos
a',
       'Iris-versicolor', 'Iris-virginica', 'Iris-versicolor',
       'Iris-virginica', 'Iris-versicolor', 'Iris-virginica',
       'Iris-virginica', 'Iris-setosa', 'Iris-versicolor', 'Iris-setos
a',
       'Iris-versicolor', 'Iris-virginica', 'Iris-virginica',
       'Iris-setosa', 'Iris-versicolor', 'Iris-virginica',
       'Iris-versicolor'], dtype=object)
```

In [13]:

```
from sklearn.metrics import confusion_matrix, accuracy_score
```

In [14]:

```
print("Confusion Matrix of Iris Dataset is : \n", confusion_matrix(y_test, y_pred))
```

```
Confusion Matrix of Iris Dataset is :
 [[14  0  0]
 [ 0 17  1]
 [ 0  1 12]]
```

In [15]:

```
print("Accuracy Score of Iris Dataset is : ", accuracy_score(y_test, y_pred))
```

```
Accuracy Score of Iris Dataset is :  0.9555555555555556
```

Accuracy of Iris Dataset using Decision Tree Classifier is 95.55%

Classification 2: KNN Classification

In [16]:

```
from sklearn.neighbors import KNeighborsClassifier
```

In [17]:

```
Knn = KNeighborsClassifier(n_neighbors=3, metric="euclidean")
```

In [18]:

```
Knn.fit(x_train,y_train)
```

Out[18]:

```
KNeighborsClassifier(metric='euclidean', n_neighbors=3)
```

In [19]:

```
knn_y_pred = Knn.predict(x_test)
knn_y_pred
```

Out[19]:

```
array(['Iris-setosa', 'Iris-versicolor', 'Iris-versicolor', 'Iris-seto
sa',
       'Iris-virginica', 'Iris-versicolor', 'Iris-virginica',
       'Iris-setosa', 'Iris-setosa', 'Iris-virginica', 'Iris-versicolo
r',
       'Iris-setosa', 'Iris-virginica', 'Iris-versicolor',
       'Iris-versicolor', 'Iris-setosa', 'Iris-versicolor',
       'Iris-versicolor', 'Iris-setosa', 'Iris-setosa', 'Iris-versicol
or',
       'Iris-versicolor', 'Iris-versicolor', 'Iris-setosa',
       'Iris-virginica', 'Iris-versicolor', 'Iris-setosa', 'Iris-setos
a',
       'Iris-versicolor', 'Iris-virginica', 'Iris-versicolor',
       'Iris-virginica', 'Iris-versicolor', 'Iris-virginica',
       'Iris-virginica', 'Iris-setosa', 'Iris-versicolor', 'Iris-setos
a',
       'Iris-versicolor', 'Iris-virginica', 'Iris-virginica',
       'Iris-setosa', 'Iris-versicolor', 'Iris-virginica',
       'Iris-versicolor'], dtype=object)
```

In [20]:

```
print("Confusion Matrix of Iris Dataset is : \n", confusion_matrix(y_test, knn_y_pre
```

```
Confusion Matrix of Iris Dataset is :
 [[14  0  0]
 [ 0 18  0]
 [ 0  1 12]]
```

In [21]:

```
print("Accuracy Score of Iris Dataset is : ", accuracy_score(y_test, knn_y_pred))
```

```
Accuracy Score of Iris Dataset is :  0.9777777777777777
```

Accuracy of Iris Dataset using KNN Classifier is 97.77%

Accuracy of Iris Dataset varies between Decision Tree

Classifier and KNN Classifier in the ratio 0.95:0.97