NAME:-Sheetal Dilip Fegade
ROLL NO:- 58
PRACTICAL NO. :-
PRACTICAL NAME :- . IMPLEMENT SIMPLE KNN USING EUCLIDEAN DISTANCE IN PYTHON.

```
from pandas import DataFrame
from sklearn.datasets import load_iris
data_b = load_iris()
df= DataFrame(data_b.data, columns=data_b.feature_names)
df['target'] = data_b.target
#print(df)
#print(data_b.DESCR)
print("Dataset Labels=",data_b.target_names)
from sklearn.neighbors import KNeighborsClassifier
from sklearn import metrics
from sklearn.metrics import confusion_matrix
from sklearn.model_selection import train_test_split
X_train, X_test, Y_train, y_test = train_test_split(df[data_b.feature_names], df['target'],
random_state=1)
print(X_train.head(6))
print(Y_train.head(6))
print(X_test.head())
clf = KNeighborsClassifier(n_neighbors=6)
clf.fit(X_train, Y_train) # model is trained
y_pred=clf.predict(X_test)
#print(y_test, y_pred)
print("Accuracy:",metrics.accuracy_score(y_test, y_pred))
cm = confusion_matrix(y_test, y_pred)
print("Confusion Matrix:")
print(cm)
```

OUTPUT:

```
Dataset Labels= ['setosa' 'versicolor' 'virginica']
    sepal length (cm)  sepal width (cm)  petal length (cm)  petal width (cm)
54                6.5               2.8                4.6               1.5
108               6.7               2.5                5.8               1.8
112               6.8               3.0                5.5               2.1
17                5.1               3.5                1.4               0.3
119               6.0               2.2                5.0               1.5
103               6.3               2.9                5.6               1.8
54    1
108   2
112   2
17    0
```

119 2

103 2

Name: target, dtype: int32

	sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)
14	5.8	4.0	1.2	0.2
98	5.1	2.5	3.0	1.1
75	6.6	3.0	4.4	1.4
16	5.4	3.9	1.3	0.4
131	7.9	3.8	6.4	2.0

Accuracy: 1.0

Confusion Matrix:

[[13 0 0]

[0 16 0]

[0 0 9]]

CODE: FOR BREAST CANCER DATA SET

```
from pandas import DataFrame
# from sklearn.datasets import load_iris
from sklearn.datasets import load_breast_cancer
from sklearn.neighbors import KNeighborsClassifier
from sklearn import metrics
from sklearn.metrics import confusion_matrix
data_b = load_breast_cancer()
df= DataFrame(data_b.data, columns=data_b.feature_names)
df['target'] = data_b.target
#print(df)
#print(data_b.DESCR)
print("Dataset Labels=",data_b.target_names)
from sklearn.model_selection import train_test_split
X_train, X_test, Y_train, y_test = train_test_split(df[data_b.feature_names], df['target'],
random_state=1)
print(X_train.head(6))
print(Y_train.head(6))
print(X_test.head())
clf = KNeighborsClassifier(n_neighbors=6)
clf.fit(X_train, Y_train) # model is trained
y_pred=clf.predict(X_test)
#print(y_test, y_pred)
print("Accuracy:",metrics.accuracy_score(y_test, y_pred))
cm = confusion_matrix(y_test, y_pred)
print("Confusion Matrix:")
print(cm)
```

OUTPUT:

Dataset Labels= ['malignant' 'benign']

	mean radius	mean texture	...	worst symmetry	worst fractal dimension
562	15.22	30.62	...	0.4089	0.14090
291	14.96	19.10	...	0.2962	0.08472
16	14.68	20.13	...	0.3029	0.08216
546	10.32	16.35	...	0.2681	0.07399
293	11.85	17.46	...	0.3101	0.07007
350	11.66	17.07	...	0.2731	0.06825

[6 rows x 30 columns]

562 0

291 1

16 0

546 1

293 1

350 1

Name: target, dtype: int32

	mean radius	mean texture	...	worst symmetry	worst fractal dimension
421	14.69	13.98	...	0.2827	0.09208
47	13.17	18.66	...	0.3900	0.11790
292	12.95	16.02	...	0.3380	0.09584
186	18.31	18.58	...	0.3206	0.06938
414	15.13	29.81	...	0.3233	0.06165

[5 rows x 30 columns]

Accuracy: 0.9370629370629371

Confusion Matrix:

[[51 4]

[5 83]]