Importing Library

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import matplotlib.axes as ax
```

About Data

This data comprised of kernels belonging to two different varieties of wheat: Kama and Canadian.

To construct the data, seven geometric parameters of wheat kernels were measured:

1. area A,
2. perimeter P,
3. compactness C = 4.pi.A / P^2,
4. length of kernel,
5. width of kernel,
6. asymmetry coefficient
7. length of kernel groove.

Loading Data

```
data = pd.read_csv('seed_data.csv')
```

A little bit of data exploration

```
data.head(10)
```

	area	perimeter	compactness	length	width	asymmetry	length_kernel_groove
0	15.26	14.84	0.8710	5.763	3.312	2.221	5.220
1	14.88	14.57	0.8811	5.554	3.333	1.018	4.956
2	14.29	14.09	0.9050	5.291	3.337	2.699	4.825
3	13.84	13.94	0.8955	5.324	3.379	2.259	4.805
4	16.14	14.99	0.9034	5.658	3.562	1.355	5.175

```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 210 entries, 0 to 209
Data columns (total 8 columns):
 #   Column                Non-Null Count  Dtype  
---  ------                --------------  -----  
 0   area                  210 non-null    float64
 1   perimeter             210 non-null    float64
 2   compactness           210 non-null    float64
 3   length                210 non-null    float64
 4   width                 210 non-null    float64
 5   asymmetry             210 non-null    float64
 6   length_kernel_groove  210 non-null    float64
 7   kernel                210 non-null    int64  
dtypes: float64(7), int64(1)
memory usage: 13.2 KB
```

Counting unique values

```
data['kernel'].unique()
```

```
array([0, 1, 2], dtype=int64)
```

```
data['kernel'].value_counts()
```

```
0    70
1    70
2    70
Name: kernel, dtype: int64
```

Splitting data

```
# training dataset and labels
x = data.drop(data.columns[[7]], axis = 1)
y = data['kernel']
```

```
x.head()
```

	area	perimeter	compactness	length	width	asymmetry	length_kernel_groove
0	15.26	14.84	0.8710	5.763	3.312	2.221	5.220
1	14.88	14.57	0.8811	5.554	3.333	1.018	4.956
2	14.29	14.09	0.9050	5.291	3.337	2.699	4.825
3	13.84	13.94	0.8955	5.324	3.379	2.259	4.805
4	16.14	14.99	0.9034	5.658	3.562	1.355	5.175

```
y.head()
```

```
0    0
1    0
2    0
3    0
4    0
Name: kernel, dtype: int64
```

```
# splitting into training and testing data
from sklearn.model_selection import train_test_split
xtrain, xtest, ytrain, ytest = train_test_split(x, y, test_size = 0.25, random_state =0)
```

Feature Scaling

```
from sklearn.preprocessing import StandardScaler
sc_x = StandardScaler()
xtrain = sc_x.fit_transform(xtrain)
xtest = sc_x.transform(xtest)
```

K Nearest Neighbor

Training model

```
# using sklearn library
# fitting k-NN model

from sklearn.neighbors import KNeighborsClassifier

knn_classifier = KNeighborsClassifier(n_neighbors = 3, p=2)
knn_classifier.fit(xtrain, ytrain)
```

```
     KNeighborsClassifier(n_neighbors=3)
```

Predicting Test_input

```
y_pred = knn_classifier.predict(xtest)
```

Results

```
# Building confusion MAtrix

from sklearn.metrics import confusion_matrix
cm = confusion_matrix(ytest, y_pred)
```

```
cm
```

```
     array([[16,  1,  0],
            [ 2, 19,  0],
            [ 0,  0, 15]], dtype=int64)
```

```
# finding accuracy

from sklearn.metrics import accuracy_score
accuracy_score(ytest, y_pred)
```

```
     0.9433962264150944
```

Keep Following and Practice more :)

GeeksforGeeks