

```
In [ ]: import pandas as pd
import numpy as np
data = pd.read_csv('C:\\Users\\huzai\\OneDrive\\Desktop\\fruit.txt', delimiter='\\t')
```

```
In [ ]: X = data[['mass', 'width', 'height', 'color_score']]
y = data['fruit_label']
```

```
In [ ]: data.head()
```

Out [ ]:

	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 [ ]: def euclidean_distance(point1, point2):
return np.sqrt(np.sum((point1 - point2) ** 2))

def predict_knn(X_train, y_train, index, k=3):
x_test = X_train.iloc[index]
distances = [euclidean_distance(x_test, X_train.iloc[i]) for i in range(len(X_train))]
k_indices = np.argsort(distances)[:k]
k_nearest_labels = [y_train.iloc[i] for i in k_indices]
most_common = np.bincount(k_nearest_labels).argmax()
return most_common
```

```
In [ ]: predictions = [predict_knn(X, y, i, k=3) for i in range(len(X))]
```

```
In [ ]: accuracy = np.mean(predictions == y.values)
print('Accuracy:', accuracy*100)
```

Accuracy: 81.35593220338984

```
In [ ]: for i in range(len(y)):
result_df = pd.DataFrame({'Actual Price': y, 'Predicted Price': predictions})
print(result_df.head(20))
```

	Actual Price	Predicted Price
0	1	1
1	1	1
2	1	1
3	2	2
4	2	2
5	2	2
6	2	2
7	2	2
8	1	1
9	1	1
10	1	1
11	1	1
12	1	3
13	1	1
14	1	1
15	1	1
16	1	1
17	1	1
18	1	1
19	1	1