

# Project - 4 (DATASET: Breast Cancer Prediction)

In [1]:

```
import pandas as pd
from matplotlib import pyplot as plt
%matplotlib inline
```

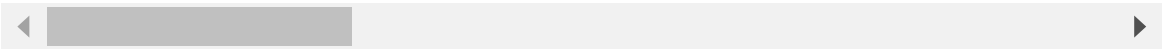
In [2]:

```
#Reading the data
df=pd.read_csv(r"C:\Users\mural\Downloads\BreastCancerPrediction.csv")
df
```

Out[2]:

	id	diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	smoothr
0	842302	M	17.99	10.38	122.80	1001.0	
1	842517	M	20.57	17.77	132.90	1326.0	
2	84300903	M	19.69	21.25	130.00	1203.0	
3	84348301	M	11.42	20.38	77.58	386.1	
4	84358402	M	20.29	14.34	135.10	1297.0	
...	...	...	...	...	...	...	
564	926424	M	21.56	22.39	142.00	1479.0	
565	926682	M	20.13	28.25	131.20	1261.0	
566	926954	M	16.60	28.08	108.30	858.1	
567	927241	M	20.60	29.33	140.10	1265.0	
568	92751	B	7.76	24.54	47.92	181.0	

569 rows × 33 columns



## DATA CLEANING AND PREPROCESSING

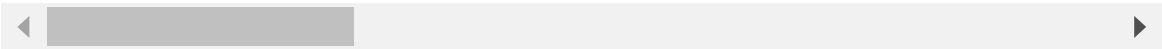
In [3]:

```
df.head()
```

Out[3]:

	id	diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	smoothness
0	842302	M	17.99	10.38	122.80	1001.0	
1	842517	M	20.57	17.77	132.90	1326.0	
2	84300903	M	19.69	21.25	130.00	1203.0	
3	84348301	M	11.42	20.38	77.58	386.1	
4	84358402	M	20.29	14.34	135.10	1297.0	

5 rows × 33 columns



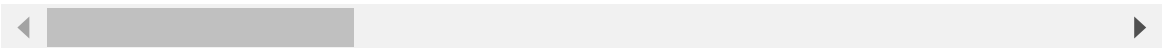
In [4]:

```
df.head()
```

Out[4]:

	id	diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	smoothness
0	842302	M	17.99	10.38	122.80	1001.0	
1	842517	M	20.57	17.77	132.90	1326.0	
2	84300903	M	19.69	21.25	130.00	1203.0	
3	84348301	M	11.42	20.38	77.58	386.1	
4	84358402	M	20.29	14.34	135.10	1297.0	

5 rows × 33 columns



In [5]:

```
#checking for null values  
df.isnull().sum()
```

Out[5]:

```
id                0  
diagnosis         0  
radius_mean      0  
texture_mean     0  
perimeter_mean   0  
area_mean        0  
smoothness_mean  0  
compactness_mean 0  
concavity_mean   0  
concave points_mean 0  
symmetry_mean    0  
fractal_dimension_mean 0  
radius_se        0  
texture_se       0  
perimeter_se     0  
area_se          0  
smoothness_se    0  
compactness_se   0  
concavity_se     0  
concave points_se 0  
symmetry_se      0  
fractal_dimension_se 0  
radius_worst     0  
texture_worst    0  
perimeter_worst  0  
area_worst       0  
smoothness_worst 0  
compactness_worst 0  
concavity_worst  0  
concave points_worst 0  
symmetry_worst   0  
fractal_dimension_worst 0  
Unnamed: 32      569  
dtype: int64
```

In [6]:

```
#Checking for duplicate values  
df.duplicated().sum()
```

Out[6]:

```
0
```

In [7]:

```
df.shape
```

Out[7]:

```
(569, 33)
```

In [8]:

```
df.describe()
```

Out[8]:

	id	radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_
count	5.690000e+02	569.000000	569.000000	569.000000	569.000000	569.0
mean	3.037183e+07	14.127292	19.289649	91.969033	654.889104	0.0
std	1.250206e+08	3.524049	4.301036	24.298981	351.914129	0.0
min	8.670000e+03	6.981000	9.710000	43.790000	143.500000	0.0
25%	8.692180e+05	11.700000	16.170000	75.170000	420.300000	0.0
50%	9.060240e+05	13.370000	18.840000	86.240000	551.100000	0.0
75%	8.813129e+06	15.780000	21.800000	104.100000	782.700000	0.1
max	9.113205e+08	28.110000	39.280000	188.500000	2501.000000	0.1

8 rows × 32 columns



In [9]:

```
df.sum()
```

Out[9]:

```
id
085
diagnosis
M...
radius_mean
429
texture_mean
5.81
perimeter_mean
0.38
In [10]:
plt.scatter(df["radius_mean"],df["texture_mean"])
plt.xlabel("radius_mean")
plt.ylabel("texture_mean")
Out[10]:
compactness_mean
002
concave points_mean
811
radius_worst
189
texture_worst
0.34
perimeter_worst
1.63
area_worst
1.8
smoothness_worst
773
compactness_worst
144.67
concave points_worst
347
symmetry_worst
```

17281572

8038.

1097

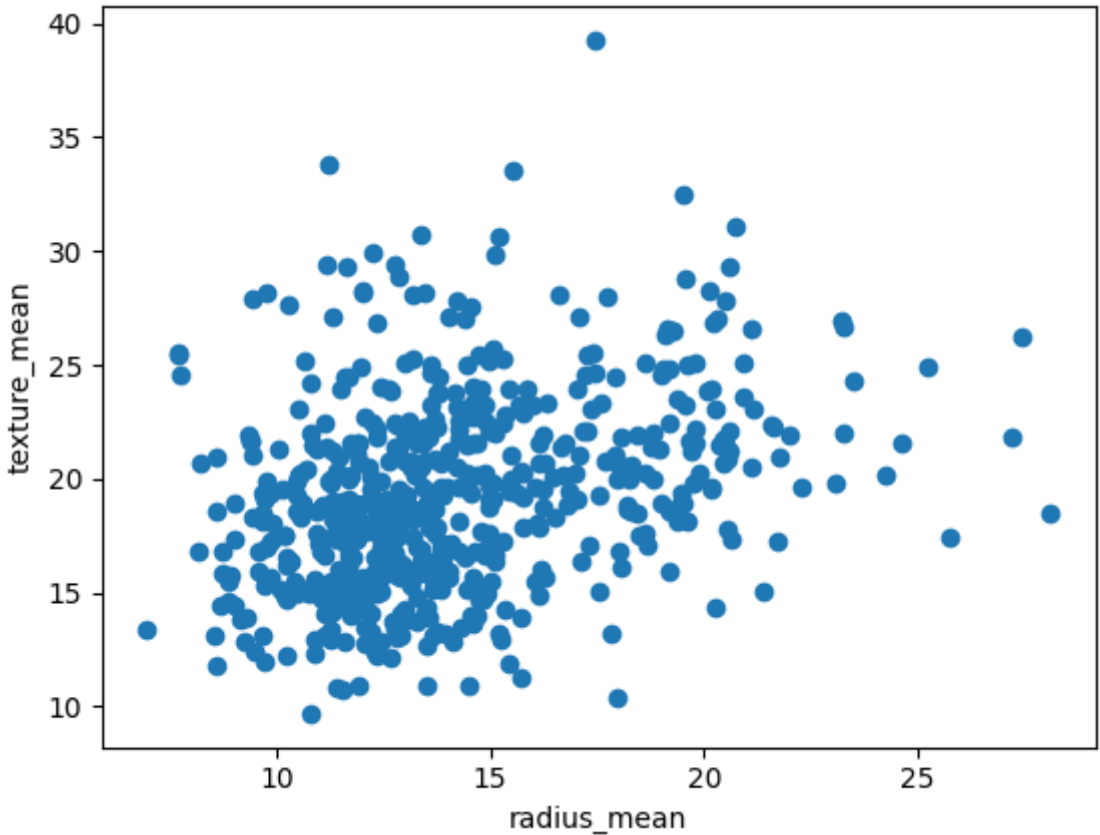
5233

37263

54.

59.37

50.526



9257.

1461

6103

50105

75.31

144.67

347

65.210

165.

In a Jupyter environment, please rerun this cell to show the HTML representation or trust the notebook.  
On GitHub, the HTML representation is unable to render, please try loading this page with nbviewer.org.

053

In [12]: Fractal\_dimension\_worst

47.76

```
517 y_predicted=km.fit_predict(df[["radius_mean","texture_mean"]])
Unamed: 32
y_predicted
0.0
```

dtype: object

C:\Users\mural\AppData\Local\Programs\Python\Python311\Lib\site-packages  
 \sklearn\cluster\\_kmeans.py:870: FutureWarning: The default value of `n\_init`  
 will change from 10 to 'auto' in 1.4. Set the value of `n\_init` explicitly  
 to suppress the warning  
 warnings.warn(

Out[12]:

```
array([3, 4, 4, 7, 4, 3, 4, 1, 5, 5, 1, 1, 2, 5, 5, 6, 1, 1, 4, 3, 3, 0,
       3, 2, 1, 3, 1, 4, 5, 3, 2, 7, 2, 2, 1, 1, 1, 7, 5, 1, 5, 5, 2, 1,
       5, 4, 7, 7, 0, 5, 5, 3, 7, 4, 1, 7, 4, 1, 7, 0, 0, 7, 5, 0, 5, 5,
       7, 7, 7, 3, 4, 0, 2, 3, 7, 1, 0, 3, 2, 7, 5, 3, 2, 2, 0, 4, 1, 2,
       5, 3, 5, 1, 3, 7, 1, 2, 7, 7, 0, 1, 5, 0, 7, 7, 7, 3, 7, 7, 4, 5,
       7, 5, 1, 7, 0, 5, 0, 3, 1, 4, 0, 4, 4, 0, 3, 3, 5, 4, 3, 2, 0, 1,
       1, 3, 4, 5, 7, 0, 3, 0, 0, 1, 7, 3, 0, 0, 7, 1, 3, 7, 5, 7, 0, 0,
       3, 7, 1, 1, 0, 0, 7, 4, 4, 5, 4, 1, 0, 1, 2, 3, 0, 1, 3, 0, 0, 0,
       7, 1, 5, 0, 4, 2, 1, 0, 1, 0, 4, 7, 7, 3, 5, 5, 7, 6, 5, 3, 5, 4,
       4, 1, 7, 1, 2, 5, 7, 3, 7, 1, 5, 3, 4, 7, 4, 2, 5, 3, 7, 7, 4, 2,
       3, 3, 7, 1, 3, 3, 0, 3, 5, 5, 1, 6, 6, 2, 0, 1, 2, 4, 6, 6, 3, 0,
       7, 5, 2, 7, 7, 0, 5, 0, 2, 7, 4, 3, 4, 3, 2, 3, 1, 6, 2, 1, 1, 1,
       1, 2, 7, 5, 3, 7, 3, 0, 4, 0, 2, 7, 0, 4, 7, 3, 2, 0, 4, 1, 3, 7,
       5, 0, 7, 7, 1, 1, 3, 7, 0, 3, 0, 7, 1, 5, 4, 7, 2, 7, 7, 5, 3, 0,
       0, 0, 7, 3, 0, 0, 7, 7, 0, 4, 7, 7, 0, 4, 0, 4, 0, 7, 3, 7, 1, 1,
       3, 7, 7, 0, 7, 1, 3, 4, 7, 2, 3, 7, 0, 4, 0, 0, 7, 3, 0, 0, 7, 1,
       4, 5, 0, 7, 7, 3, 0, 7, 7, 5, 7, 1, 3, 4, 2, 7, 4, 4, 1, 3, 4, 4,
       3, 3, 7, 6, 3, 7, 0, 0, 5, 7, 3, 5, 0, 3, 0, 2, 0, 7, 1, 4, 7, 3,
       7, 7, 0, 7, 4, 0, 7, 3, 0, 7, 3, 5, 4, 7, 7, 7, 5, 1, 6, 5, 5, 1,
       0, 5, 7, 3, 0, 1, 7, 5, 0, 5, 7, 7, 1, 7, 4, 4, 3, 1, 7, 3, 1, 3,
       7, 2, 3, 7, 4, 5, 2, 3, 1, 4, 5, 2, 6, 3, 7, 6, 6, 5, 5, 6, 2, 2,
       6, 7, 7, 1, 1, 7, 2, 7, 7, 6, 3, 6, 0, 3, 1, 3, 0, 1, 7, 1, 3, 3,
       3, 3, 3, 4, 7, 1, 5, 3, 4, 0, 1, 1, 7, 7, 4, 4, 3, 5, 3, 4, 0, 0,
       7, 7, 3, 5, 0, 3, 1, 3, 1, 7, 4, 4, 7, 3, 0, 4, 7, 7, 0, 0, 7, 0,
       3, 0, 7, 7, 3, 4, 7, 4, 5, 5, 5, 5, 0, 5, 5, 6, 1, 5, 7, 7, 7, 5,
       5, 5, 6, 5, 6, 6, 7, 6, 5, 5, 6, 6, 6, 2, 4, 2, 6, 2, 5])
```

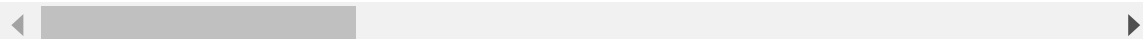
In [13]:

```
df["cluster"]=y_predicted
df.head()
```

Out[13]:

	id	diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	smoothnes
0	842302	M	17.99	10.38	122.80	1001.0	
1	842517	M	20.57	17.77	132.90	1326.0	
2	84300903	M	19.69	21.25	130.00	1203.0	
3	84348301	M	11.42	20.38	77.58	386.1	
4	84358402	M	20.29	14.34	135.10	1297.0	

5 rows × 34 columns



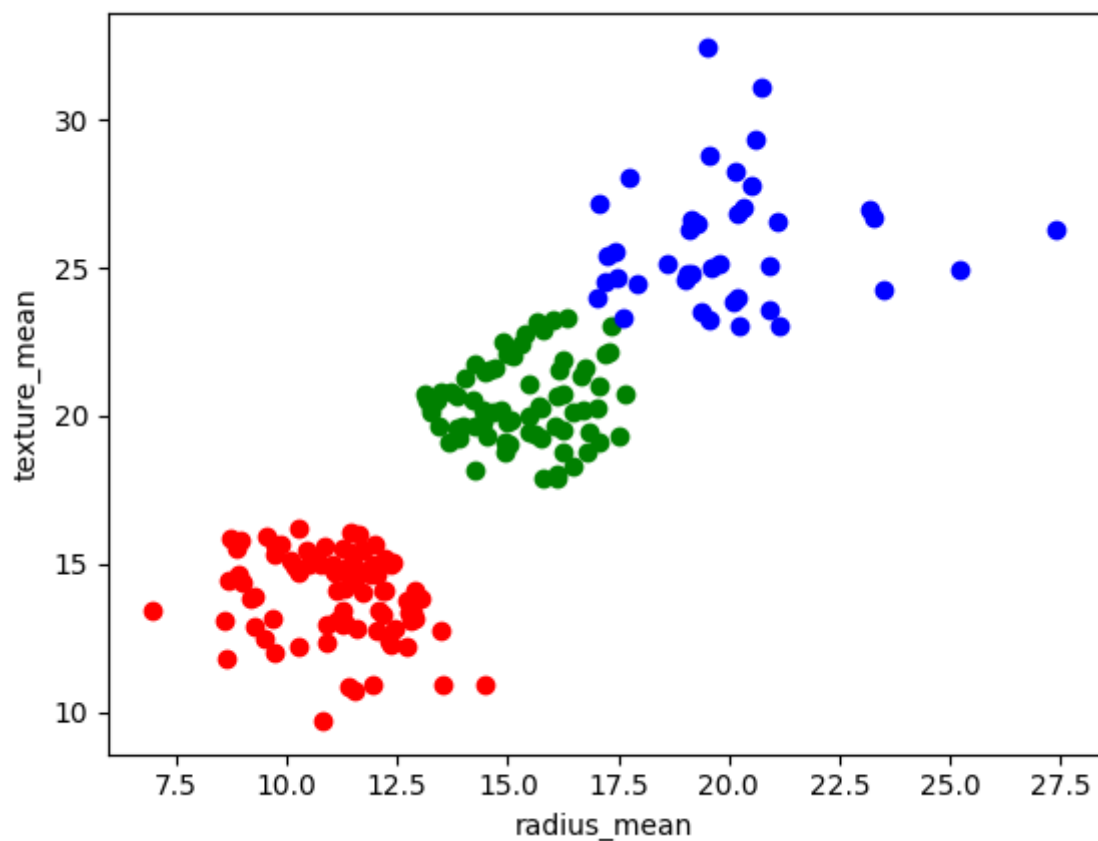


In [14]:

```
df1=df[df.cluster==0]
df2=df[df.cluster==1]
df3=df[df.cluster==2]
plt.scatter(df1["radius_mean"],df1["texture_mean"],color="red")
plt.scatter(df2["radius_mean"],df2["texture_mean"],color="green")
plt.scatter(df3["radius_mean"],df3["texture_mean"],color="blue")
plt.xlabel("radius_mean")
plt.ylabel("texture_mean")
```

Out[14]:

Text(0, 0.5, 'texture\_mean')



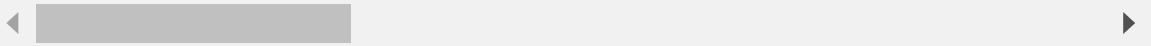
In [15]:

```
from sklearn.preprocessing import MinMaxScaler
scaler=MinMaxScaler()
scaler.fit(df[["texture_mean"]])
df["texture_mean"]=scaler.transform(df[["texture_mean"]])
df.head()
```

Out[15]:

	id	diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	smoothness
0	842302	M	17.99	0.022658	122.80	1001.0	
1	842517	M	20.57	0.272574	132.90	1326.0	
2	84300903	M	19.69	0.390260	130.00	1203.0	
3	84348301	M	11.42	0.360839	77.58	386.1	
4	84358402	M	20.29	0.156578	135.10	1297.0	

5 rows × 34 columns



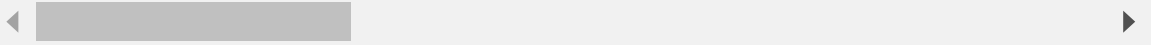
In [16]:

```
scaler.fit(df[["radius_mean"]])
df["radius_mean"]=scaler.transform(df[["radius_mean"]])
df.head()
```

Out[16]:

	id	diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	smoothness
0	842302	M	0.521037	0.022658	122.80	1001.0	
1	842517	M	0.643144	0.272574	132.90	1326.0	
2	84300903	M	0.601496	0.390260	130.00	1203.0	
3	84348301	M	0.210090	0.360839	77.58	386.1	
4	84358402	M	0.629893	0.156578	135.10	1297.0	

5 rows × 34 columns



In [17]:

```
y_predicted=km.fit_predict(df[["radius_mean","texture_mean"]])
y_predicted
```

C:\Users\mural\AppData\Local\Programs\Python\Python311\Lib\site-packages  
 \sklearn\cluster\\_kmeans.py:870: FutureWarning: The default value of `n\_i  
 nit` will change from 10 to 'auto' in 1.4. Set the value of `n\_init` expl  
 icitly to suppress the warning  
 warnings.warn(

Out[17]:

```
array([0, 4, 4, 2, 4, 0, 4, 5, 5, 3, 5, 0, 6, 5, 5, 3, 5, 5, 4, 0, 0, 7,
       0, 1, 5, 4, 5, 4, 5, 4, 6, 2, 6, 6, 0, 5, 5, 2, 5, 5, 5, 2, 6, 5,
       5, 4, 7, 2, 7, 5, 2, 0, 2, 4, 5, 2, 4, 5, 2, 7, 7, 2, 5, 7, 3, 5,
       2, 2, 2, 0, 4, 7, 6, 0, 2, 5, 0, 4, 6, 2, 2, 0, 1, 6, 7, 4, 5, 6,
       5, 0, 5, 5, 0, 2, 5, 6, 2, 2, 7, 5, 3, 7, 2, 2, 2, 0, 2, 2, 1, 2,
       2, 2, 5, 2, 7, 2, 7, 0, 5, 4, 7, 4, 1, 0, 0, 0, 3, 4, 0, 6, 7, 5,
       5, 0, 4, 5, 2, 7, 0, 7, 7, 0, 2, 0, 7, 7, 2, 5, 0, 0, 5, 2, 7, 7,
       0, 2, 4, 4, 7, 7, 2, 4, 4, 5, 1, 5, 7, 4, 6, 0, 7, 5, 0, 7, 7, 7,
       2, 5, 5, 0, 1, 6, 5, 7, 5, 7, 4, 2, 2, 0, 5, 5, 2, 3, 5, 0, 5, 4,
       4, 5, 2, 4, 1, 5, 2, 0, 2, 4, 5, 0, 4, 2, 1, 6, 5, 0, 2, 2, 4, 6,
       0, 0, 2, 5, 0, 0, 7, 0, 3, 5, 4, 3, 3, 6, 7, 5, 1, 4, 3, 6, 0, 0,
       2, 5, 6, 2, 0, 0, 3, 7, 6, 2, 4, 4, 4, 0, 6, 0, 5, 3, 6, 6, 4, 5,
       4, 6, 2, 5, 0, 2, 0, 7, 1, 7, 6, 2, 7, 4, 0, 0, 6, 7, 4, 5, 0, 2,
       2, 0, 2, 2, 5, 5, 0, 2, 0, 0, 7, 2, 0, 2, 4, 2, 6, 2, 2, 3, 0, 7,
       0, 0, 2, 0, 0, 7, 2, 2, 7, 4, 2, 2, 7, 4, 0, 4, 7, 2, 0, 2, 5, 5,
       0, 2, 2, 7, 2, 4, 0, 4, 2, 1, 0, 7, 7, 4, 7, 7, 2, 0, 7, 7, 2, 5,
       1, 3, 7, 2, 2, 0, 7, 2, 2, 5, 2, 4, 0, 4, 6, 2, 4, 1, 5, 0, 4, 4,
       0, 0, 2, 3, 0, 2, 7, 7, 5, 2, 0, 5, 7, 0, 7, 6, 7, 7, 5, 1, 2, 0,
       5, 2, 7, 2, 4, 7, 2, 0, 7, 2, 0, 5, 4, 2, 2, 2, 2, 5, 3, 2, 2, 5,
       7, 2, 2, 0, 7, 5, 2, 2, 7, 2, 2, 2, 5, 2, 4, 4, 0, 5, 2, 0, 5, 0,
       2, 6, 0, 2, 4, 3, 6, 0, 5, 4, 2, 6, 3, 0, 2, 3, 3, 3, 3, 3, 6, 1,
       3, 2, 2, 5, 5, 2, 6, 2, 2, 3, 0, 3, 7, 0, 5, 0, 7, 5, 2, 5, 0, 0,
       0, 0, 0, 4, 7, 4, 5, 0, 4, 7, 5, 5, 2, 2, 4, 4, 0, 3, 0, 1, 7, 7,
       2, 2, 0, 5, 7, 0, 5, 0, 5, 2, 4, 4, 2, 0, 7, 1, 2, 5, 7, 7, 2, 7,
       0, 7, 2, 2, 0, 4, 2, 4, 5, 3, 3, 3, 7, 3, 3, 3, 5, 5, 7, 7, 2, 3,
       2, 2, 3, 2, 3, 3, 2, 3, 5, 3, 3, 3, 3, 6, 1, 6, 6, 6, 3])
```

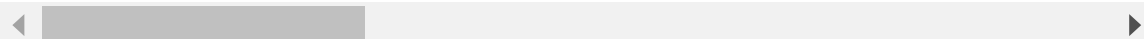
In [18]:

```
df["New Cluster"]=y_predicted
df.head()
```

Out[18]:

	id	diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	smoothnes
0	842302	M	0.521037	0.022658	122.80	1001.0	
1	842517	M	0.643144	0.272574	132.90	1326.0	
2	84300903	M	0.601496	0.390260	130.00	1203.0	
3	84348301	M	0.210090	0.360839	77.58	386.1	
4	84358402	M	0.629893	0.156578	135.10	1297.0	

5 rows × 35 columns

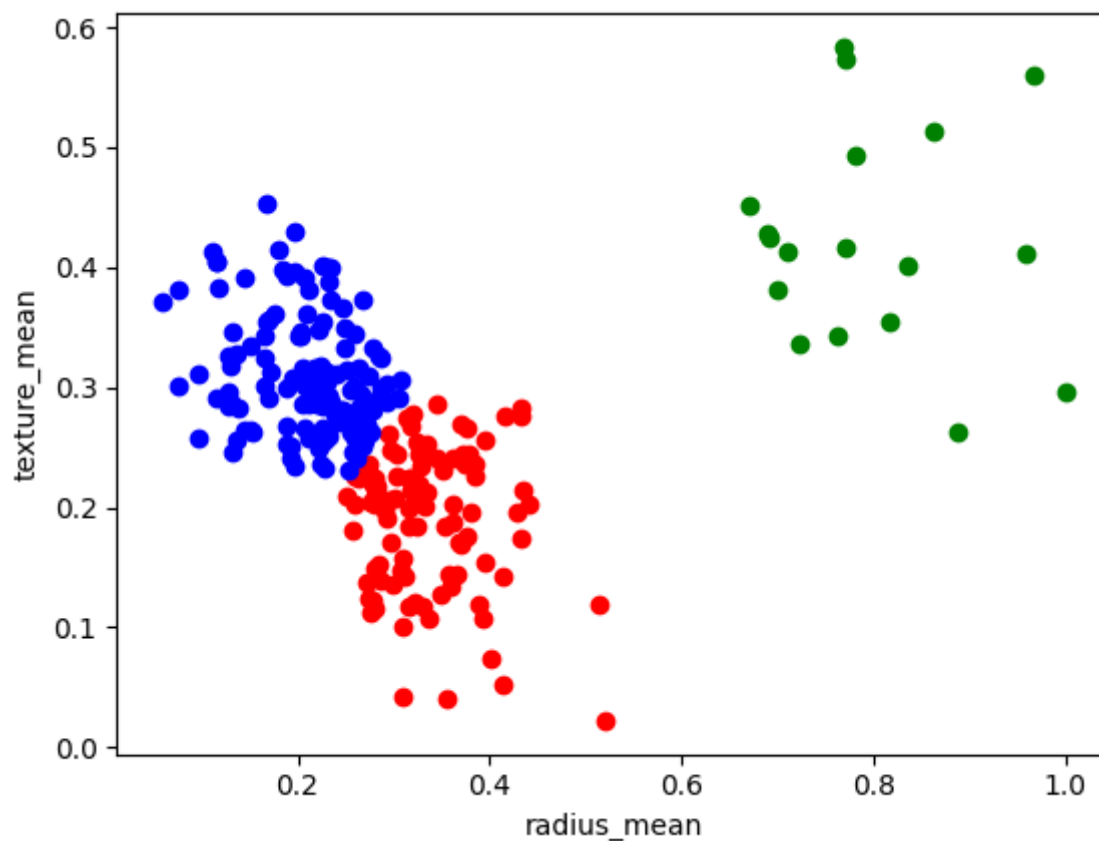


In [19]:

```
df1=df[df["New Cluster"]==0]
df2=df[df["New Cluster"]==1]
df3=df[df["New Cluster"]==2]
plt.scatter(df1["radius_mean"],df1["texture_mean"],color="red")
plt.scatter(df2["radius_mean"],df2["texture_mean"],color="green")
plt.scatter(df3["radius_mean"],df3["texture_mean"],color="blue")
plt.xlabel("radius_mean")
plt.ylabel("texture_mean")
```

Out[19]:

Text(0, 0.5, 'texture\_mean')

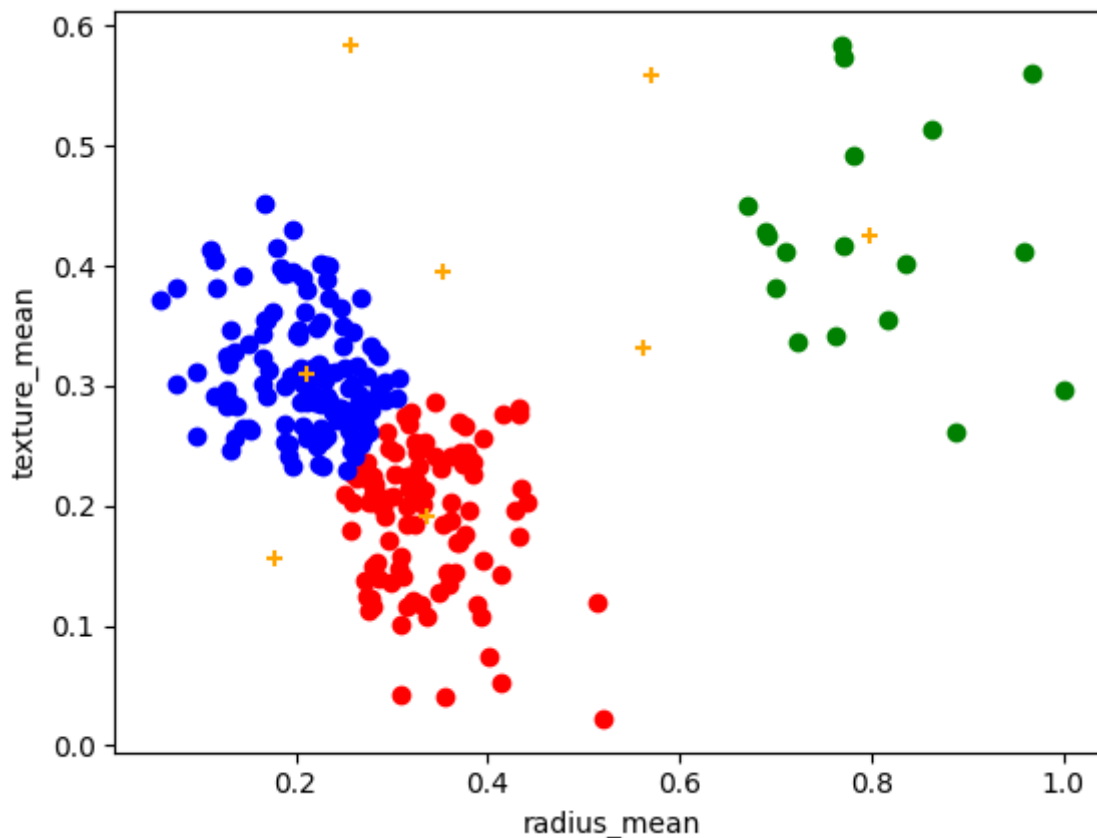


In [20]:

```
df1=df[df["New Cluster"]==0]
df2=df[df["New Cluster"]==1]
df3=df[df["New Cluster"]==2]
plt.scatter(df1["radius_mean"],df1["texture_mean"],color="red")
plt.scatter(df2["radius_mean"],df2["texture_mean"],color="green")
plt.scatter(df3["radius_mean"],df3["texture_mean"],color="blue")
plt.scatter(km.cluster_centers_[0],km.cluster_centers_[1],color="orange",marker="+")
plt.xlabel("radius_mean")
plt.ylabel("texture_mean")
```

Out[20]:

Text(0, 0.5, 'texture\_mean')



In [22]:

```
k_rng=range(1,10)
sse=[]
```

In [23]:

```

for k in k_rng:
    km=KMeans(n_clusters=k)
    km.fit(df[["radius_mean", "texture_mean"]])
    sse.append(km.inertia_)
#km.inertia_ will give you the value of sum of square error
print(sse)
plt.plot(k_rng, sse)
plt.xlabel("K")
plt.ylabel("Sum of Squared Error")

```

C:\Users\mural\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\cluster\\_kmeans.py:870: FutureWarning: The default value of `n\_init` will change from 10 to 'auto' in 1.4. Set the value of `n\_init` explicitly to suppress the warning

```
warnings.warn(
```

C:\Users\mural\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\cluster\\_kmeans.py:870: FutureWarning: The default value of `n\_init` will change from 10 to 'auto' in 1.4. Set the value of `n\_init` explicitly to suppress the warning

```
warnings.warn(
```

C:\Users\mural\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\cluster\\_kmeans.py:870: FutureWarning: The default value of `n\_init` will change from 10 to 'auto' in 1.4. Set the value of `n\_init` explicitly to suppress the warning

```
warnings.warn(
```

C:\Users\mural\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\cluster\\_kmeans.py:870: FutureWarning: The default value of `n\_init` will change from 10 to 'auto' in 1.4. Set the value of `n\_init` explicitly to suppress the warning

```
warnings.warn(
```

C:\Users\mural\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\cluster\\_kmeans.py:870: FutureWarning: The default value of `n\_init` will change from 10 to 'auto' in 1.4. Set the value of `n\_init` explicitly to suppress the warning

```
warnings.warn(
```

C:\Users\mural\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\cluster\\_kmeans.py:870: FutureWarning: The default value of `n\_init` will change from 10 to 'auto' in 1.4. Set the value of `n\_init` explicitly to suppress the warning

```
warnings.warn(
```

C:\Users\mural\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\cluster\\_kmeans.py:870: FutureWarning: The default value of `n\_init` will change from 10 to 'auto' in 1.4. Set the value of `n\_init` explicitly to suppress the warning

```
warnings.warn(
```

C:\Users\mural\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\cluster\\_kmeans.py:870: FutureWarning: The default value of `n\_init` will change from 10 to 'auto' in 1.4. Set the value of `n\_init` explicitly to suppress the warning

```
warnings.warn(
```

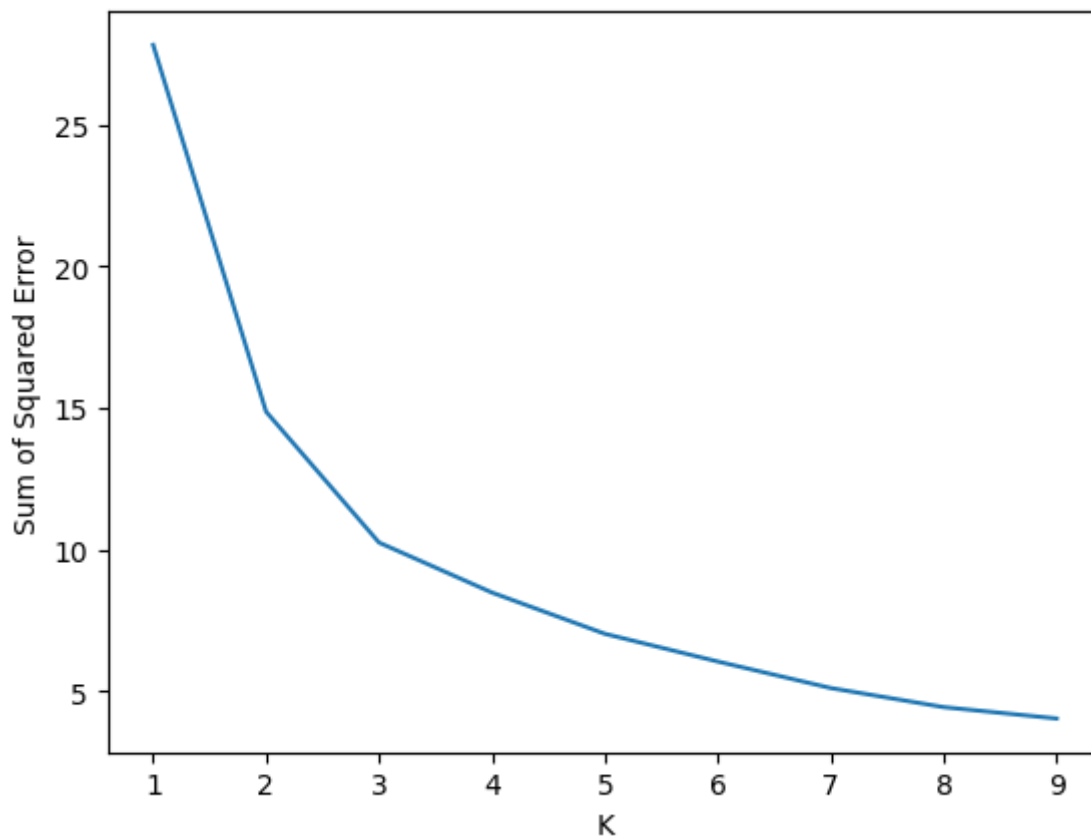
C:\Users\mural\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\cluster\\_kmeans.py:870: FutureWarning: The default value of `n\_init` will change from 10 to 'auto' in 1.4. Set the value of `n\_init` explicitly to suppress the warning

```
warnings.warn(
```

```
[27.81750759504308, 14.87203295827117, 10.2527514961052, 8.48752519907283  
4, 7.031082156151861, 6.052732460629174, 5.116507804943426, 4.44269210256  
1712, 4.04124880202688]
```

Out[23]:

Text(0, 0.5, 'Sum of Squared Error')



#### CONCLUSION

for the given dataset we can use multiple models, for that models we get different types that accuracies is not good so, that's why we will take it as a clustering and done with