In [1]:

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

In [2]:

```
df=pd.read_csv(r"C:\Users\magam\Downloads\BreastCancerPrediction.csv")
df
```

Out[2]:

	id	diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	smoo
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

In [3]:

```
df.describe()
```

Out[3]:

	id	radius_mean	texture_mean	perimeter_mean	area_mean	smoothne
count	5.690000e+02	569.000000	569.000000	569.000000	569.000000	56
mean	3.037183e+07	14.127292	19.289649	91.969033	654.889104	
std	1.250206e+08	3.524049	4.301036	24.298981	351.914129	
min	8.670000e+03	6.981000	9.710000	43.790000	143.500000	
25%	8.692180e+05	11.700000	16.170000	75.170000	420.300000	
50%	9.060240e+05	13.370000	18.840000	86.240000	551.100000	
75%	8.813129e+06	15.780000	21.800000	104.100000	782.700000	
max	9.113205e+08	28.110000	39.280000	188.500000	2501.000000	

8 rows × 32 columns

In [4]:

```
df.columns
```

Out[4]:

```
Index(['id', 'diagnosis', 'radius_mean', 'texture_mean', 'perimeter_mea
n',
       'area_mean', 'smoothness_mean', 'compactness_mean', 'concavity_m
ean',
       'concave points_mean', 'symmetry_mean', 'fractal_dimension_mea
n',
       'radius_se', 'texture_se', 'perimeter_se', 'area_se', 'smoothnes
s_se',
       'compactness_se', 'concavity_se', 'concave points_se', 'symmetry
_se',
       'fractal_dimension_se', 'radius_worst', 'texture_worst',
       'perimeter_worst', 'area_worst', 'smoothness_worst',
       'compactness_worst', 'concavity_worst', 'concave points_worst',
       'symmetry_worst', 'fractal_dimension_worst', 'Unnamed: 32'],
      dtype='object')
```

In [5]:

```
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]:

```
del df["Unnamed: 32"]
```

In [7]:

```
df
```

Out[7]:

	id	diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	smooth
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	

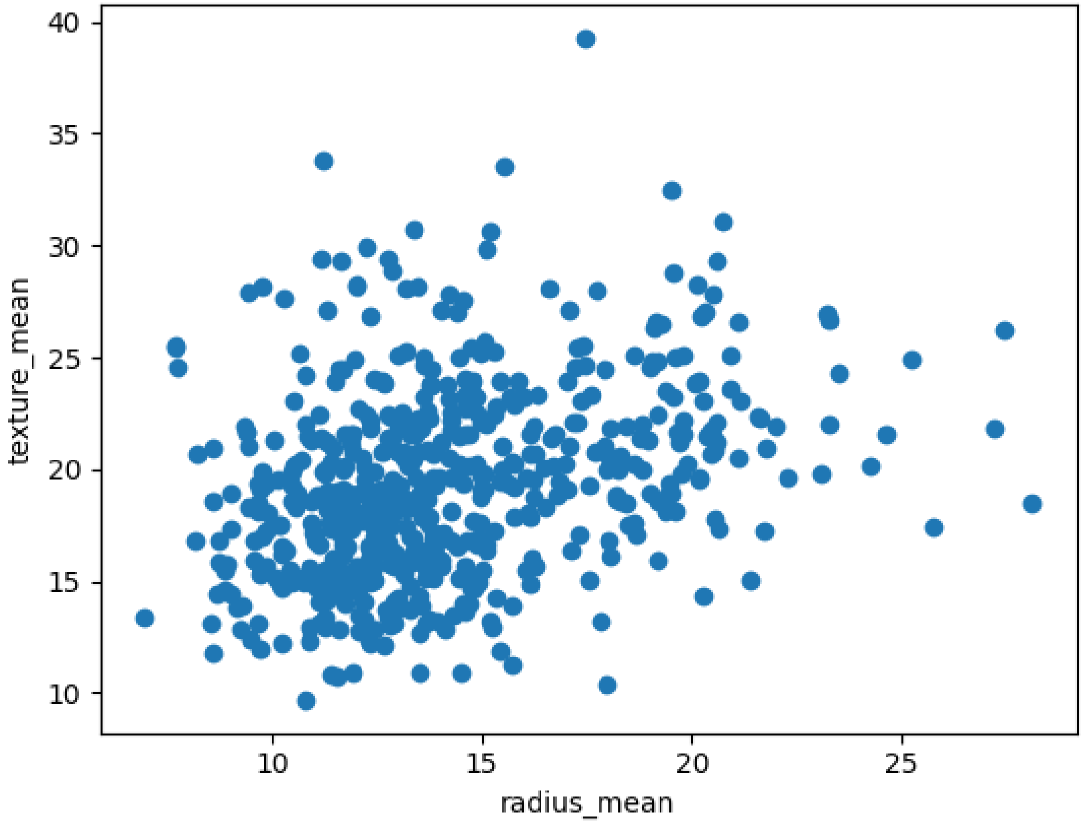
569 rows × 32 columns

In [8]:

```
plt.scatter(df["radius_mean"],df["texture_mean"])
plt.xlabel("radius_mean")
plt.ylabel("texture_mean")
```

Out[8]:

```
Text(0, 0.5, 'texture_mean')
```



In [9]:

```
from sklearn.cluster import KMeans
km=KMeans()
km
```

Out[9]:

```
▾ KMeans
KMeans()
```

In [10]:

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

```
C:\Users\magam\AppData\Local\Programs\Python\Python311\Lib\site-package
s\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[10]:

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

In [11]:

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

Out[11]:

	id	diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	smooth
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	

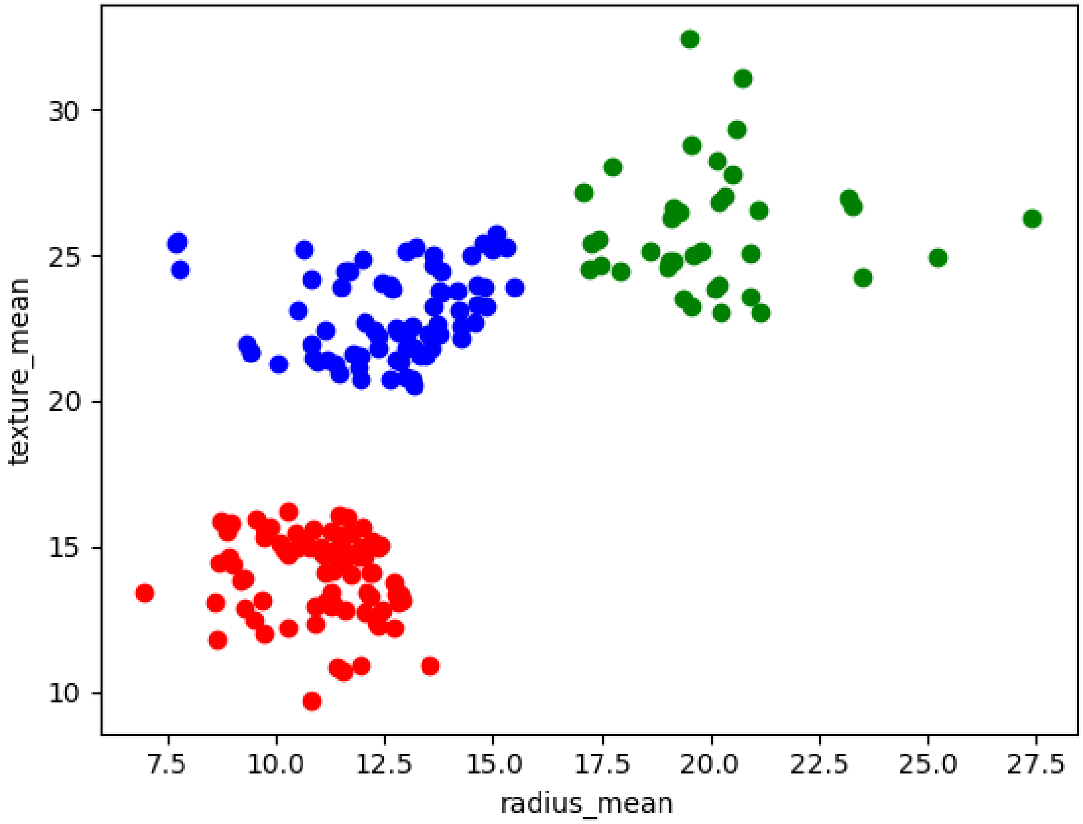
5 rows × 33 columns

In [12]:

```
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[12]:

```
Text(0, 0.5, 'texture_mean')
```



In [13]:

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

Out[13]:

	id	diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	smooth
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 × 33 columns

In [14]:

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

Out[14]:

	id	diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	smooth
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 × 33 columns

In [15]:

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

```
C:\Users\magam\AppData\Local\Programs\Python\Python311\Lib\site-package
s\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[15]:

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

In [16]:

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

Out[16]:

	id	diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	smooth
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	

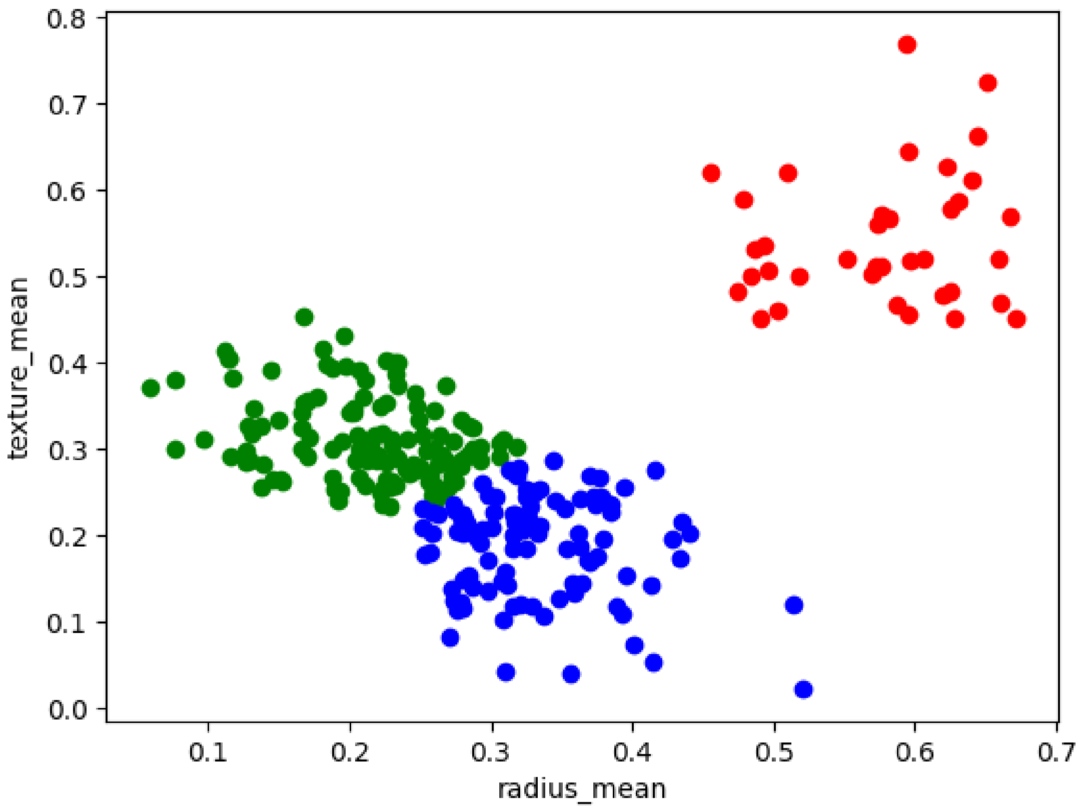
5 rows × 34 columns

In [17]:

```
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[17]:

```
Text(0, 0.5, 'texture_mean')
```



In [18]:

```
km.cluster_centers_
```

Out[18]:

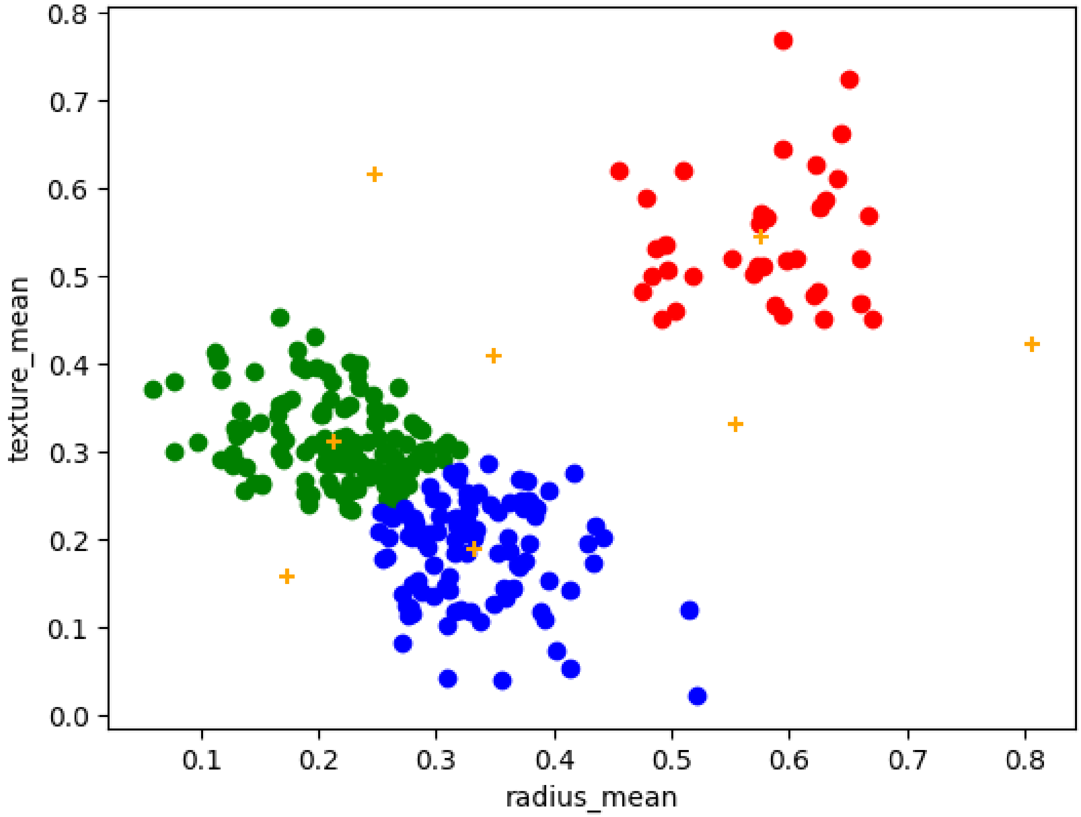
```
array([[0.57605341, 0.54408687],
       [0.21276186, 0.31200594],
       [0.33158632, 0.18812362],
       [0.80589822, 0.42316338],
       [0.17405343, 0.15815861],
       [0.55409393, 0.33027383],
       [0.24753115, 0.61622301],
       [0.34814903, 0.40844623]])
```

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.scatter(km.cluster_centers_[:,0],km.cluster_centers_[:,1],color="orange",mark
plt.xlabel("radius_mean")
plt.ylabel("texture_mean")
```

Out[19]:

```
Text(0, 0.5, 'texture_mean')
```



In [20]:

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

In [21]:

```
for k in k_rng:
 km=KMeans(n_clusters=k)
 km.fit(df[["radius_mean","texture_mean"]])
 sse.append(km.inertia_)

print(sse)
plt.plot(k_rng,sse)
plt.xlabel("K")
plt.ylabel("Sum of Squared Error")
```

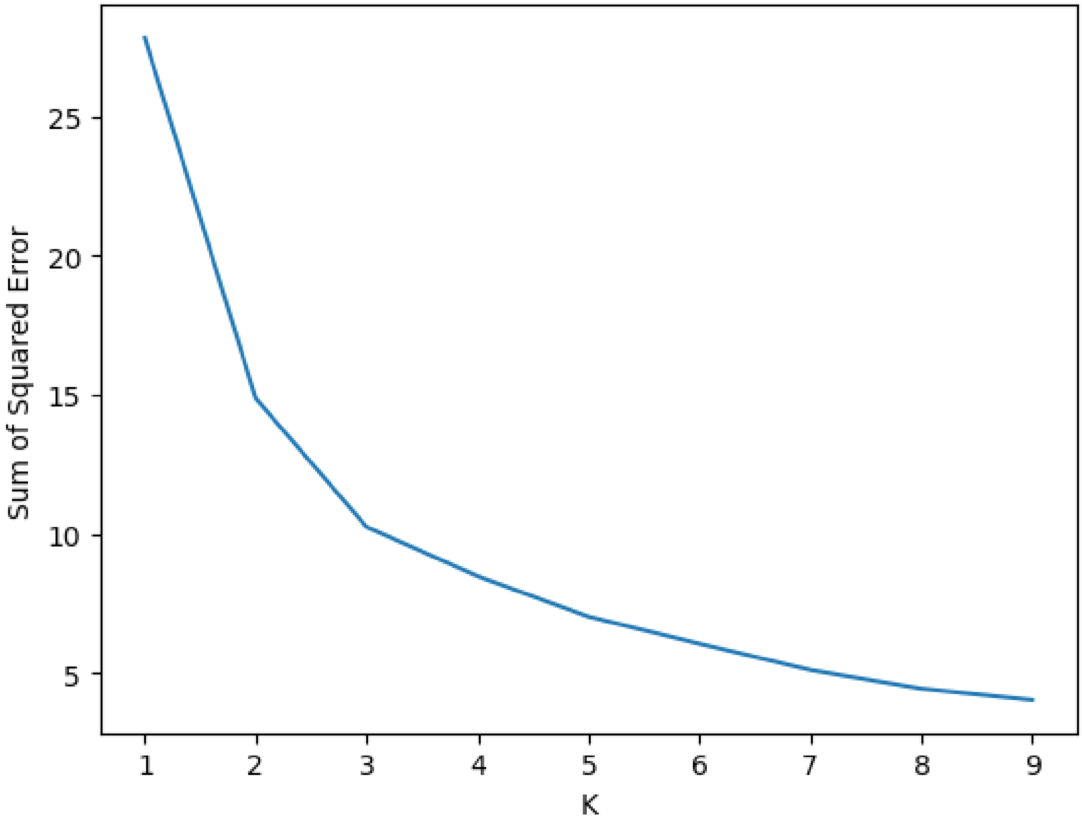
```
C:\Users\magam\AppData\Local\Programs\Python\Python311\Lib\site-package
s\sklearn\cluster\_kmeans.py:870: FutureWarning: The default value of `
n_init` will change from 10 to 'auto' in 1.4. Set the value of `n_init`
explicitly to suppress the warning
  warnings.warn(
C:\Users\magam\AppData\Local\Programs\Python\Python311\Lib\site-package
s\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\magam\AppData\Local\Programs\Python\Python311\Lib\site-package
s\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\magam\AppData\Local\Programs\Python\Python311\Lib\site-package
s\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\magam\AppData\Local\Programs\Python\Python311\Lib\site-package
s\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\magam\AppData\Local\Programs\Python\Python311\Lib\site-package
s\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, 8.484725277027
605, 7.029817500713495, 6.058842607216886, 5.117927753802226, 4.4444615
30393398, 4.038561172431516]
```

```
C:\Users\magam\AppData\Local\Programs\Python\Python311\Lib\site-package
s\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\magam\AppData\Local\Programs\Python\Python311\Lib\site-package
s\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\magam\AppData\Local\Programs\Python\Python311\Lib\site-package
s\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[21]:

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



done with this.