Project - 4 (DATASET: Breast Cancer Prediction)

In [21]:
```
import pandas as pd
from matplotlib import pyplot as plt
%matplotlib inline
```

In [52]:
```
df=pd.read_csv(r"C:\Users\mouni\Downloads\BreastCancerPrediction.csv")
df
```

Out[52]:

	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	

In [53]:
```
df.head()
```

Out[53]:

	id	diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_n
0	842302	M	17.99	10.38	122.80	1001.0	0.1
1	842517	M	20.57	17.77	132.90	1326.0	0.0
2	84300903	M	19.69	21.25	130.00	1203.0	0.1
3	84348301	M	11.42	20.38	77.58	386.1	0.1
4	84358402	M	20.29	14.34	135.10	1297.0	0.1

5 rows × 33 columns

In [54]:
```
df.tail()
```

Out[54]:

	id	diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_n
564	926424	M	21.56	22.39	142.00	1479.0	0.1
565	926682	M	20.13	28.25	131.20	1261.0	0.0
566	926954	M	16.60	28.08	108.30	858.1	0.0
567	927241	M	20.60	29.33	140.10	1265.0	0.1
568	92751	B	7.76	24.54	47.92	181.0	0.0

5 rows × 33 columns

In [55]:
```
df.drop(['Unnamed: 32'],axis=1)
```

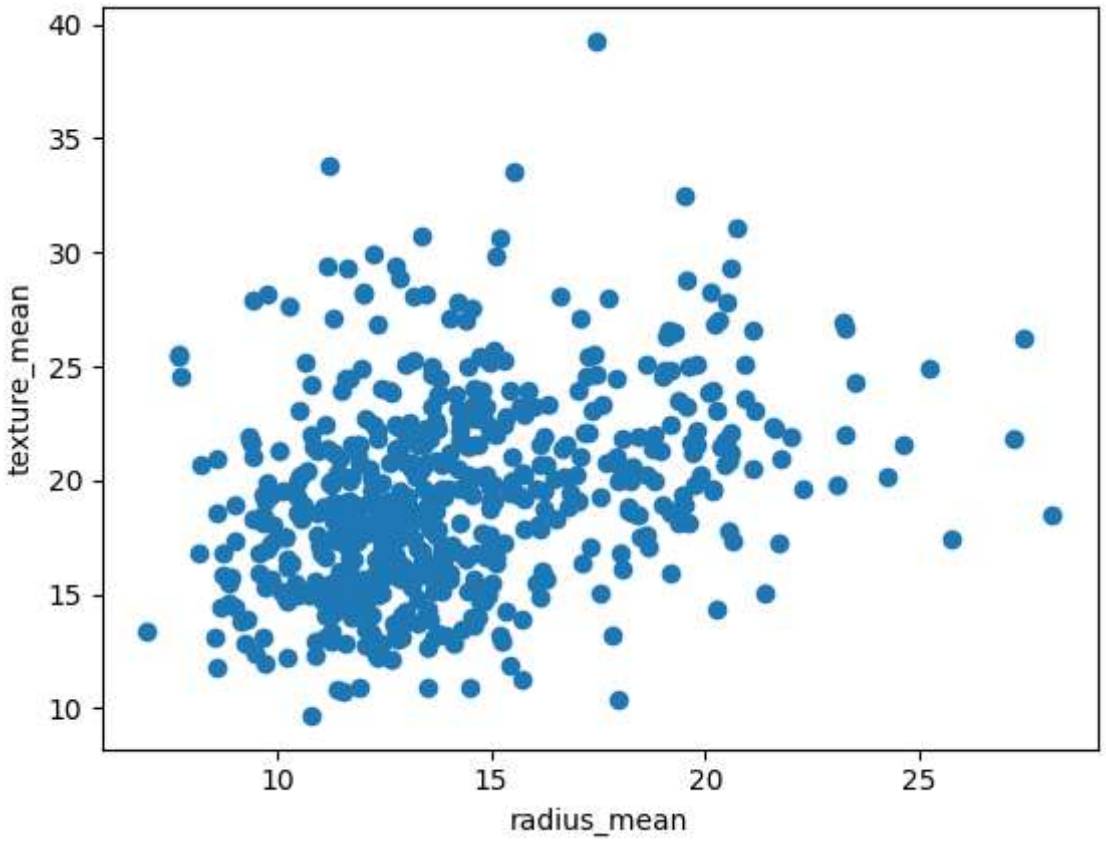
Out[55]:

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

569 rows × 32 columns

In [56]:
```
plt.scatter(df["radius_mean"],df["texture_mean"])
plt.xlabel("radius_mean")
plt.ylabel("texture_mean")
```

Out[56]:
```
Text(0, 0.5, 'texture_mean')
```



In [43]:
```
from sklearn.cluster import KMeans
km=KMeans()
km
```

Out[43]:
```
KMeans
KMeans()
```

In [57]:
```
y_predicted=km.fit_predict(df[["radius_mean","texture_mean"]])
y_predicted
```

```
C:\Users\mouni\AppData\Local\Programs\Python\Python310\lib\site-packages\skl
earn\cluster\_kmeans.py:870: FutureWarning: The default value of `n_init` wi
ll change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly to
suppress the warning
  warnings.warn(
```

Out[57]:
```
array([6, 2, 2, 4, 2, 6, 2, 7, 3, 3, 7, 7, 5, 7, 3, 0, 7, 7, 2, 6, 6, 1,
       6, 5, 7, 6, 7, 2, 3, 6, 5, 4, 7, 5, 7, 7, 7, 4, 3, 7, 3, 3, 5, 7,
       3, 2, 4, 4, 1, 3, 3, 6, 4, 2, 7, 4, 2, 7, 4, 1, 1, 4, 3, 1, 3, 3,
       4, 4, 4, 6, 2, 1, 5, 6, 4, 7, 1, 6, 5, 4, 3, 6, 5, 5, 1, 2, 7, 5,
       3, 6, 3, 7, 6, 4, 7, 5, 4, 4, 1, 7, 3, 1, 4, 4, 4, 6, 4, 4, 2, 3,
       4, 3, 7, 4, 1, 3, 1, 6, 7, 7, 1, 2, 2, 6, 6, 6, 3, 2, 6, 5, 1, 7,
       7, 6, 2, 3, 4, 1, 6, 1, 1, 7, 4, 6, 1, 1, 4, 7, 6, 4, 3, 4, 1, 1,
       6, 4, 7, 7, 1, 1, 4, 2, 2, 3, 2, 7, 1, 7, 5, 6, 1, 4, 6, 1, 1, 1,
       4, 7, 3, 1, 2, 5, 7, 1, 7, 1, 2, 4, 4, 6, 3, 3, 4, 0, 3, 6, 3, 7,
       2, 7, 4, 7, 5, 3, 4, 6, 4, 7, 3, 6, 2, 4, 2, 5, 3, 6, 4, 4, 2, 5,
       6, 6, 4, 7, 6, 6, 1, 6, 3, 3, 7, 0, 0, 5, 1, 7, 5, 2, 0, 0, 6, 1,
       4, 3, 5, 4, 4, 6, 3, 1, 5, 4, 2, 6, 2, 6, 5, 6, 7, 0, 5, 7, 7, 7,
       7, 5, 4, 3, 6, 4, 6, 1, 2, 1, 5, 4, 1, 2, 4, 6, 5, 1, 2, 7, 6, 4,
       3, 1, 4, 4, 7, 7, 6, 4, 1, 6, 1, 4, 4, 3, 2, 4, 5, 4, 4, 3, 6, 1,
       6, 6, 4, 6, 1, 1, 4, 4, 1, 2, 4, 4, 1, 2, 1, 2, 1, 4, 6, 4, 7, 7,
       6, 4, 4, 1, 4, 7, 6, 2, 4, 5, 6, 4, 1, 2, 1, 1, 4, 6, 1, 1, 4, 7,
       2, 3, 1, 4, 4, 6, 1, 4, 4, 3, 4, 7, 6, 2, 5, 4, 2, 2, 7, 6, 2, 2,
       6, 6, 4, 0, 6, 4, 1, 1, 3, 4, 6, 3, 1, 6, 1, 5, 1, 4, 7, 2, 4, 6,
       4, 4, 1, 4, 7, 1, 4, 6, 1, 4, 6, 3, 7, 4, 4, 4, 3, 7, 0, 3, 3, 7,
       1, 3, 4, 6, 1, 4, 4, 3, 1, 3, 4, 4, 7, 4, 2, 2, 6, 7, 4, 6, 7, 6,
       4, 5, 6, 4, 2, 3, 5, 6, 7, 2, 3, 5, 0, 6, 4, 0, 0, 3, 3, 0, 5, 5,
       0, 4, 4, 4, 3, 4, 7, 4, 4, 0, 6, 0, 1, 6, 7, 6, 1, 7, 4, 7, 6, 4,
       6, 4, 6, 2, 4, 7, 3, 6, 7, 1, 3, 7, 4, 4, 2, 2, 6, 3, 6, 2, 1, 1,
       4, 4, 6, 3, 1, 6, 7, 6, 7, 4, 2, 2, 4, 4, 1, 2, 4, 4, 1, 1, 4, 1,
       6, 1, 4, 4, 6, 2, 4, 2, 3, 3, 3, 3, 1, 3, 3, 0, 7, 3, 4, 4, 4, 3,
       3, 3, 0, 3, 0, 0, 4, 0, 3, 3, 0, 0, 0, 5, 2, 5, 0, 5, 3])
```

In [45]:
```
s["cluster"]=y_predicted
s.head()
```

Out[45]:

	id	diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_n
0	842302	M	0.521037	10.38	122.80	1001.0	0.1
1	842517	M	0.643144	17.77	132.90	1326.0	0.0
2	84300903	M	0.601496	21.25	130.00	1203.0	0.1
3	84348301	M	0.210090	20.38	77.58	386.1	0.1
4	84358402	M	0.629893	14.34	135.10	1297.0	0.1

5 rows × 34 columns

In [58]:
```
from sklearn.preprocessing import MinMaxScaler
scaler=MinMaxScaler()
scaler.fit(df[["texture_mean"]])
df["texture_mean"]=scaler.transform(df[["texture_mean"]])
df.head()
```

Out[58]:

	id	diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_n
0	842302	M	17.99	0.022658	122.80	1001.0	0.1
1	842517	M	20.57	0.272574	132.90	1326.0	0.0
2	84300903	M	19.69	0.390260	130.00	1203.0	0.1
3	84348301	M	11.42	0.360839	77.58	386.1	0.1
4	84358402	M	20.29	0.156578	135.10	1297.0	0.1

5 rows × 33 columns

In [59]:
```
scaler.fit(df[["radius_mean"]])
df["radius_mean"]=scaler.transform(df[["radius_mean"]])
df.head()
```

Out[59]:

	id	diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_n
0	842302	M	0.521037	0.022658	122.80	1001.0	0.1
1	842517	M	0.643144	0.272574	132.90	1326.0	0.0
2	84300903	M	0.601496	0.390260	130.00	1203.0	0.1
3	84348301	M	0.210090	0.360839	77.58	386.1	0.1
4	84358402	M	0.629893	0.156578	135.10	1297.0	0.1

5 rows × 33 columns

In [48]:
```
y_predicted=km.fit_predict(s[["radius_mean","texture_mean"]])
y_predicted
```

```
C:\Users\mouni\AppData\Local\Programs\Python\Python310\lib\site-packages\skl
earn\cluster\_kmeans.py:870: FutureWarning: The default value of `n_init` wi
ll change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly to
suppress the warning
  warnings.warn(
```

Out[48]:
```
array([1, 3, 0, 7, 6, 6, 7, 0, 0, 4, 4, 3, 4, 4, 0, 2, 7, 0, 0, 6, 6, 1,
       6, 4, 0, 6, 0, 7, 4, 6, 4, 7, 4, 2, 3, 0, 0, 3, 4, 0, 0, 0, 4, 7,
       0, 3, 3, 7, 6, 0, 0, 6, 3, 7, 0, 7, 7, 0, 7, 1, 6, 0, 0, 1, 4, 4,
       0, 7, 3, 3, 0, 6, 4, 6, 3, 7, 1, 6, 4, 3, 0, 6, 4, 2, 6, 7, 0, 4,
       0, 6, 4, 0, 6, 3, 7, 4, 3, 7, 1, 7, 4, 1, 0, 7, 7, 6, 3, 7, 7, 0,
       3, 0, 7, 7, 6, 0, 6, 3, 0, 7, 1, 3, 7, 1, 6, 3, 4, 7, 6, 4, 1, 7,
       0, 1, 0, 0, 3, 6, 3, 1, 1, 3, 3, 6, 6, 6, 3, 7, 6, 3, 0, 0, 6, 1,
       6, 3, 0, 7, 1, 1, 7, 6, 3, 0, 0, 7, 1, 7, 4, 3, 1, 7, 1, 6, 6, 6,
       3, 7, 0, 1, 0, 2, 7, 6, 0, 6, 7, 3, 3, 6, 4, 0, 3, 2, 4, 6, 0, 0,
       0, 7, 7, 7, 2, 4, 7, 3, 3, 7, 0, 1, 0, 7, 7, 4, 4, 3, 3, 3, 0, 5,
       1, 1, 3, 7, 3, 1, 6, 6, 4, 0, 7, 2, 5, 2, 6, 0, 2, 0, 2, 5, 6, 6,
       3, 4, 4, 7, 3, 1, 4, 6, 4, 7, 7, 3, 7, 3, 2, 3, 4, 5, 2, 4, 0, 7,
       0, 5, 7, 0, 6, 7, 3, 1, 0, 6, 4, 3, 1, 7, 3, 6, 2, 1, 3, 7, 6, 3,
       0, 1, 7, 7, 7, 7, 6, 3, 1, 1, 1, 3, 3, 4, 7, 7, 4, 7, 3, 4, 6, 6,
       1, 1, 7, 6, 1, 1, 7, 3, 1, 7, 7, 3, 6, 7, 1, 0, 6, 3, 1, 3, 0, 0,
       6, 7, 7, 6, 7, 0, 6, 0, 3, 4, 3, 3, 6, 0, 6, 6, 7, 6, 6, 6, 3, 7,
       3, 4, 1, 7, 7, 6, 6, 3, 3, 0, 7, 3, 3, 0, 2, 3, 3, 0, 4, 1, 6, 3,
       6, 6, 7, 2, 6, 7, 1, 6, 0, 3, 1, 4, 1, 6, 6, 4, 1, 3, 7, 0, 3, 3,
       7, 3, 6, 3, 0, 1, 3, 6, 6, 7, 6, 0, 0, 3, 3, 3, 0, 0, 2, 0, 0, 0,
       1, 0, 7, 1, 6, 7, 7, 0, 6, 0, 3, 3, 0, 3, 7, 0, 3, 7, 7, 6, 7, 6,
       3, 4, 6, 3, 3, 4, 2, 3, 7, 0, 0, 4, 2, 1, 3, 5, 2, 4, 4, 2, 2, 2,
       2, 3, 3, 7, 0, 3, 4, 3, 7, 2, 6, 2, 6, 6, 0, 3, 6, 7, 3, 7, 1, 3,
       1, 6, 3, 7, 6, 7, 0, 1, 0, 1, 0, 7, 3, 3, 3, 0, 3, 4, 6, 7, 1, 1,
       7, 3, 6, 4, 6, 6, 0, 1, 7, 7, 0, 7, 3, 3, 1, 0, 7, 7, 6, 1, 7, 1,
       1, 1, 3, 7, 6, 0, 3, 0, 0, 4, 4, 4, 6, 4, 4, 2, 0, 4, 6, 3, 7, 4,
       0, 0, 2, 0, 2, 2, 7, 2, 0, 4, 2, 2, 5, 4, 0, 2, 2, 2, 4])
```

In [60]:
```
df["New Cluster"]=y_predicted
df.head()
```

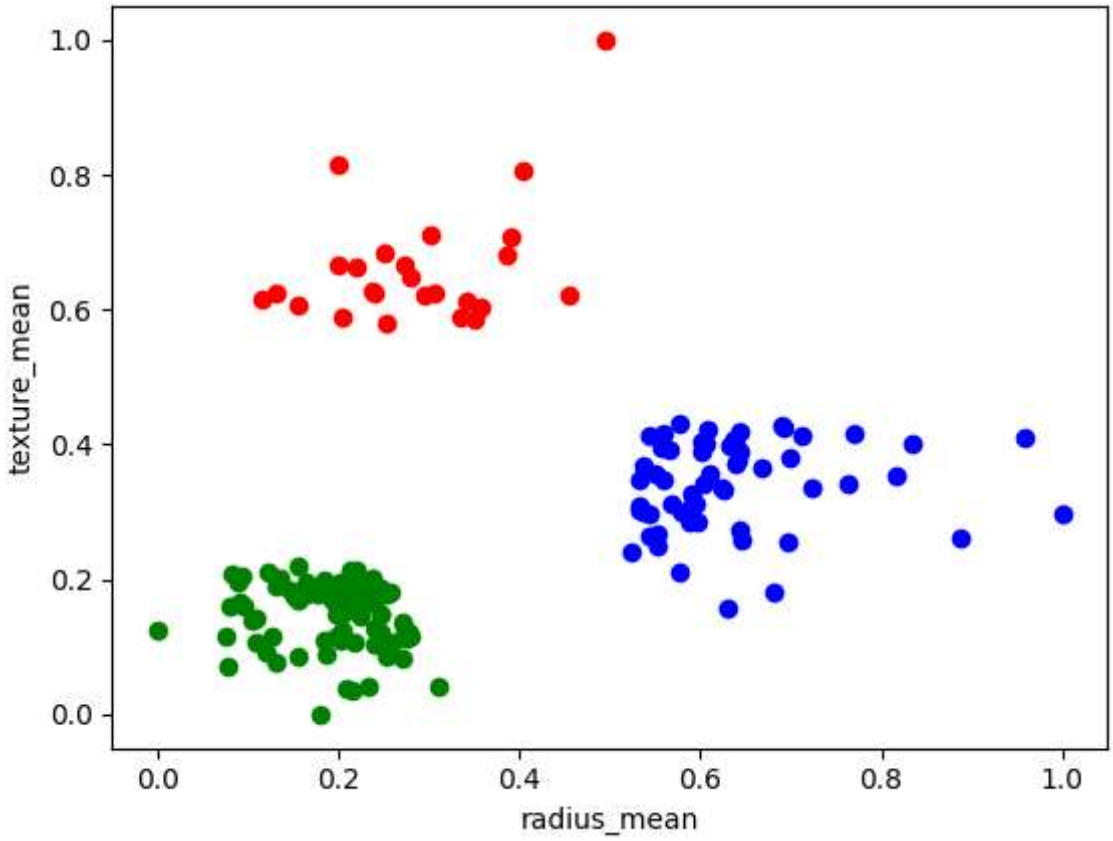
Out[60]:

	id	diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_n
0	842302	M	0.521037	0.022658	122.80	1001.0	0.1
1	842517	M	0.643144	0.272574	132.90	1326.0	0.0
2	84300903	M	0.601496	0.390260	130.00	1203.0	0.1
3	84348301	M	0.210090	0.360839	77.58	386.1	0.1
4	84358402	M	0.629893	0.156578	135.10	1297.0	0.1

5 rows × 34 columns

In [61]:
```
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[61]:
```
Text(0, 0.5, 'texture_mean')
```

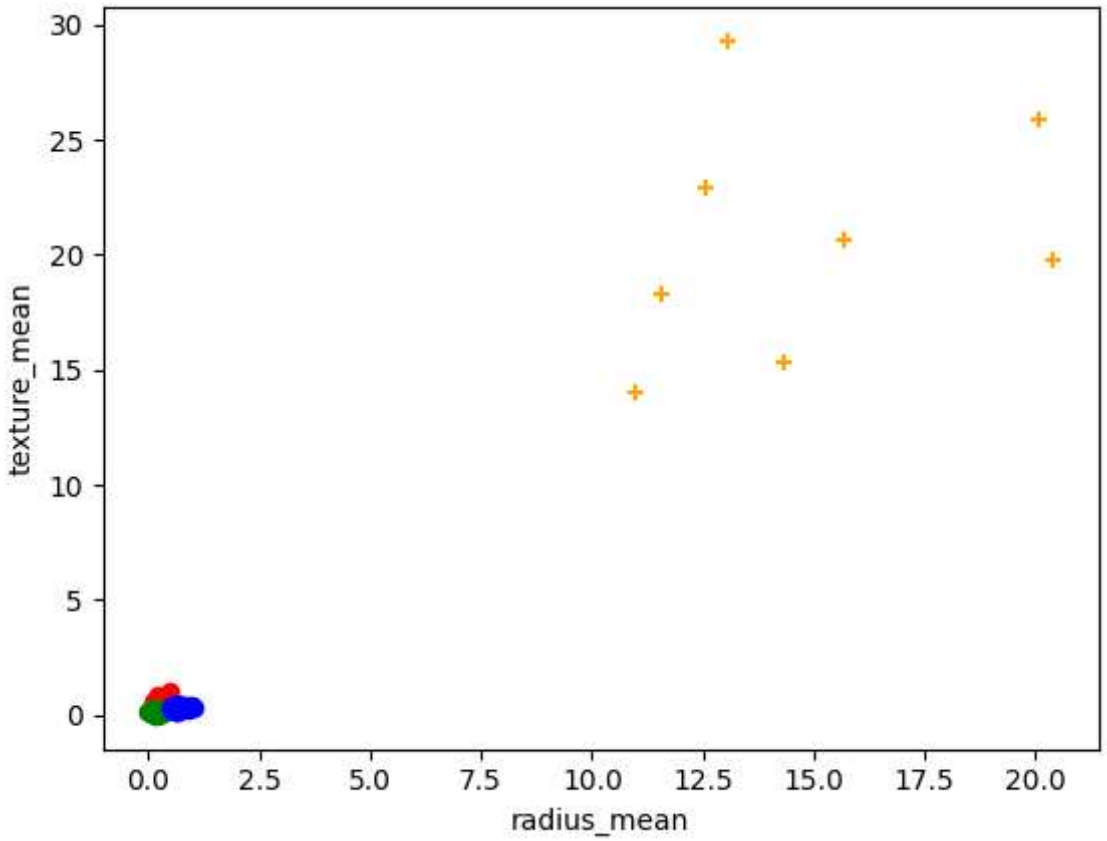


In [62]:
```
km.cluster_centers_
```

Out[62]:
```
array([[13.05352   , 29.3064    ],
       [10.9873    , 14.001625  ],
       [20.4062069 , 19.77706897],
       [12.56168919, 22.92527027],
       [11.58201538, 18.338     ],
       [20.07157895, 25.92605263],
       [14.31370787, 15.29797753],
       [15.7024    , 20.65106667]])
```

In [63]:
```
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",m
plt.xlabel("radius_mean")
plt.ylabel("texture_mean")
```

Out[63]:
```
Text(0, 0.5, 'texture_mean')
```



In [64]:
```
k_rng=range(1,10)
sse=[]
```

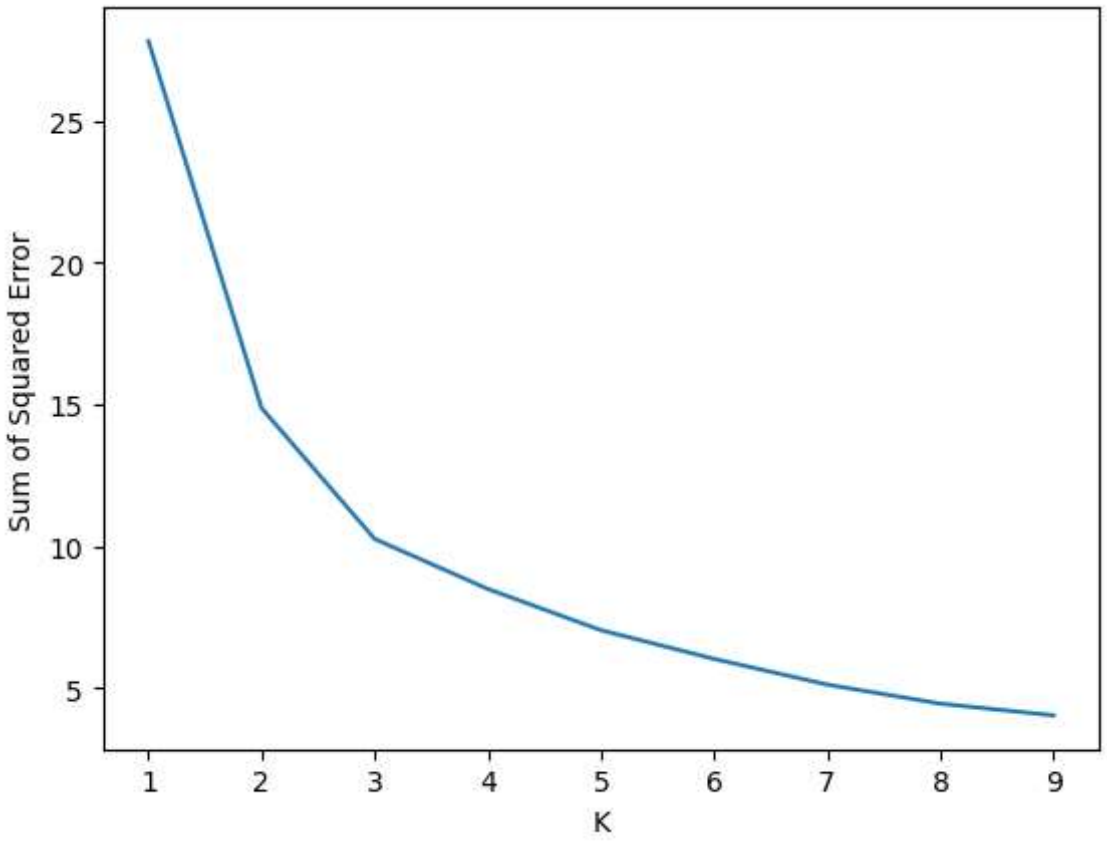
In [65]:
```python
for k in k_rng:
    km=KMeans(n_clusters=k)
    km.fit(df[["radius_mean","texture_mean"]])
    sse.append(km.inertia_)
print(sse)
plt.plot(k_rng,sse)
plt.xlabel("K")
plt.ylabel("Sum of Squared Error")
```

```
C:\Users\mouni\AppData\Local\Programs\Python\Python310\lib\site-packages\skl
earn\cluster\_kmeans.py:870: FutureWarning: The default value of `n_init` wi
ll change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly to
suppress the warning
  warnings.warn(
C:\Users\mouni\AppData\Local\Programs\Python\Python310\lib\site-packages\skl
earn\cluster\_kmeans.py:870: FutureWarning: The default value of `n_init` wi
ll change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly to
suppress the warning
  warnings.warn(
C:\Users\mouni\AppData\Local\Programs\Python\Python310\lib\site-packages\skl
earn\cluster\_kmeans.py:870: FutureWarning: The default value of `n_init` wi
ll change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly to
suppress the warning
  warnings.warn(
C:\Users\mouni\AppData\Local\Programs\Python\Python310\lib\site-packages\skl
earn\cluster\_kmeans.py:870: FutureWarning: The default value of `n_init` wi
ll change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly to
suppress the warning
  warnings.warn(
C:\Users\mouni\AppData\Local\Programs\Python\Python310\lib\site-packages\skl
earn\cluster\_kmeans.py:870: FutureWarning: The default value of `n_init` wi
ll change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly to
suppress the warning
  warnings.warn(
C:\Users\mouni\AppData\Local\Programs\Python\Python310\lib\site-packages\skl
earn\cluster\_kmeans.py:870: FutureWarning: The default value of `n_init` wi
ll change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly to
suppress the warning
  warnings.warn(
C:\Users\mouni\AppData\Local\Programs\Python\Python310\lib\site-packages\skl
earn\cluster\_kmeans.py:870: FutureWarning: The default value of `n_init` wi
ll change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly to
suppress the warning
  warnings.warn(
C:\Users\mouni\AppData\Local\Programs\Python\Python310\lib\site-packages\skl
earn\cluster\_kmeans.py:870: FutureWarning: The default value of `n_init` wi
ll change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly to
suppress the warning
  warnings.warn(

[27.81750759504307, 14.872032958271172, 10.252751496105196, 8.48945818213025
3, 7.035500433198194, 6.024074219955782, 5.116896853150587, 4.44439527370827
9, 4.031802888033525]
```

```
C:\Users\mouni\AppData\Local\Programs\Python\Python310\lib\site-packages\skl
earn\cluster\_kmeans.py:870: FutureWarning: The default value of `n_init` wi
ll change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly to
suppress the warning
  warnings.warn(
```

Out[65]: 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 of accuracies but that #accuracies is not good so,that's why we will take it as a clustering and done with K-Means Clustering

In [ ]: