Classification using breast cancer dataset

```
!gdown 1qwVr-_-_Ax
!unzip breast-cancer.zip
!cp -r /content/breast-cancer-wisconsin/*.csv /content
!rm -r breast-cancer-wisconsin
!rm breast-cancer.zip
!ls
```

```
Archive:  breast-cancer.zip
inflating: breast-cancer-wisconsin/breast-cancer-wisconsin.names
inflating: breast-cancer-wisconsin/breast-cancer-wisconsin.csv
breast-cancer-wisconsin.csv  sample_data
```

```
df = pd.read_csv('breast-cancer-wisconsin.csv')
df.head()
```

	id number	diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_mean	compactness_mean	concavity_mean	concave points_mean	symmetry_mean	fractal_dimension_mean	radius_se	texture_se	perimeter_se	area_se	smoothness_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
0	842302	M	17.99	10.38	122.80	1001.0	0.11840	0.27760	0.3001	0.14710	0.2419	0.07871	1.095	0.9053	8.589	153.40	0.006399	0.04904	0.05373	0.01587	0.03003	0.006193	25.38	17.33	184.60	2019.0	0.16622	0.56708	0.7119	0.2654	0.4601	0.11890
1	842517	M	20.57	17.77	132.90	1326.0	0.08474	0.07864	0.0869	0.07017	0.1812	0.05667	0.5435	0.7339	3.398	74.08	0.005225	0.01308	0.0186	0.0134	0.01389	0.003532	24.99	23.41	158.80	1956.0	0.12380	0.18660	0.2416	0.1860	0.2750	0.08902
2	84300903	M	19.69	21.25	130.00	1203.0	0.10960	0.15990	0.1974	0.12790	0.2069	0.05999	0.7456	0.7869	4.585	94.03	0.006150	0.04006	0.03832	0.02058	0.02250	0.004571	23.57	25.53	152.50	1709.0	0.14440	0.42450	0.4504	0.2243	0.3613	0.08758
3	84348301	M	11.42	20.38	77.58	386.1	0.14250	0.28390	0.2414	0.10520	0.2597	0.09744	0.4956	1.156	3.445	27.23	0.009110	0.07458	0.05661	0.01867	0.05963	0.009208	14.91	26.50	98.87	567.7	0.20980	0.86663	0.6869	0.2575	0.6638	0.17300
4	84358402	M	20.29	14.34	135.10	1297.0	0.10030	0.13280	0.1980	0.10430	0.1809	0.05883	0.7572	0.7813	5.438	94.44	0.011490	0.02461	0.05688	0.01885	0.01756	0.005115	22.54	16.67	152.20	1575.0	0.13740	0.20500	0.4000	0.1625	0.2364	0.07678

```
df.shape
```

```
(569, 33)
```

```
df.dtypes
```

```
id                           int64
diagnosis                   object
radius_mean                  float64
texture_mean                 float64
perimeter_mean               float64
area_mean                    float64
smoothness_mean              float64
compactness_mean             float64
concavity_mean               float64
concave points_mean          float64
symmetry_mean                float64
fractal_dimension_mean       float64
radius_se                    float64
texture_se                   float64
perimeter_se                 float64
area_se                      float64
smoothness_se                float64
compactness_se               float64
concavity_se                 float64
concave points_se            float64
symmetry_se                  float64
fractal_dimension_se         float64
radius_worst                 float64
texture_worst                float64
perimeter_worst              float64
area_worst                   float64
smoothness_worst             float64
compactness_worst            float64
concavity_worst              float64
concave points_worst         float64
symmetry_worst               float64
fractal_dimension_worst      float64
Unnamed: 32                  float64
dtype: object
```

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

	diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_mean	compactness_mean	concavity_mean	concave points_mean	symmetry_mean	fractal_dimension_mean	radius_se	texture_se	perimeter_se	area_se	smoothness_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
0	M	17.99	10.38	122.80	1001.0	0.11840	0.27760	0.3001	0.14710	0.2419	0.07871	1.095	0.9053	8.589	153.40	0.006399	0.04904	0.05373	0.01587	0.03003	0.006193	25.38	17.33	184.60	2019.0	0.16622	0.56708	0.7119	0.2654	0.4601	0.11890
1	M	20.57	17.77	132.90	1326.0	0.08474	0.07864	0.0869	0.07017	0.1812	0.05667	0.5435	0.7339	3.398	74.08	0.005225	0.01308	0.0186	0.0134	0.01389	0.003532	24.99	23.41	158.80	1956.0	0.12380	0.18660	0.2416	0.1860	0.2750	0.08902
2	M	19.69	21.25	130.00	1203.0	0.10960	0.15990	0.1974	0.12790	0.2069	0.05999	0.7456	0.7869	4.585	94.03	0.006150	0.04006	0.03832	0.02058	0.02250	0.004571	23.57	25.53	152.50	1709.0	0.14440	0.42450	0.4504	0.2243	0.3613	0.08758
3	M	11.42	20.38	77.58	386.1	0.14250	0.28390	0.2414	0.10520	0.2597	0.09744	0.4956	1.156	3.445	27.23	0.009110	0.07458	0.05661	0.01867	0.05963	0.009208	14.91	26.50	98.87	567.7	0.20980	0.86663	0.6869	0.2575	0.6638	0.17300
4	M	20.29	14.34	135.10	1297.0	0.10030	0.13280	0.1980	0.10430	0.1809	0.05883	0.7572	0.7813	5.438	94.44	0.011490	0.02461	0.05688	0.01885	0.01756	0.005115	22.54	16.67	152.20	1575.0	0.13740	0.20500	0.4000	0.1625	0.2364	0.07678

```
df.dtypes
```

```
diagnosis                   object
radius_mean                  float64
texture_mean                 float64
perimeter_mean               float64
area_mean                    float64
smoothness_mean              float64
compactness_mean             float64
concavity_mean               float64
concave points_mean          float64
symmetry_mean                float64
fractal_dimension_mean       float64
radius_se                    float64
texture_se                   float64
perimeter_se                 float64
area_se                      float64
smoothness_se                float64
compactness_se               float64
concavity_se                 float64
concave points_se            float64
symmetry_se                  float64
fractal_dimension_se         float64
radius_worst                 float64
texture_worst                float64
perimeter_worst              float64
area_worst                   float64
smoothness_worst             float64
compactness_worst            float64
concavity_worst              float64
concave points_worst         float64
symmetry_worst               float64
fractal_dimension_worst      float64
dtype: object
```

```
df.drop(['id number', 'Unnamed: 32'], axis=1).describe
```

```
<bound_method_NDFrame.describe of       radius_mean  texture_mean  perimeter_mean  area_mean  smoothness_mean  compactness_mean  concavity_mean  concave points_mean  symmetry_mean  fractal_dimension_mean  radius_se  texture_se  perimeter_se  area_se  smoothness_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
0         17.99         10.38          122.80     1001.0          0.11840           0.27760          0.3001              0.14710         0.2419                 0.07871      1.095      0.9053         8.589    153.40       0.006399        0.04904       0.05373            0.01587       0.03003                0.006193         25.38          17.33           184.60      2019.0            0.16622            0.56708           0.7119                0.2654          0.4601                 0.11890
1         20.57         17.77          132.90     1326.0          0.08474           0.07864          0.0869              0.07017         0.1812                 0.05667      0.5435      0.7339         3.398     74.08       0.005225        0.01308        0.0186            0.0134       0.01389                0.003532         24.99          23.41           158.80      1956.0            0.12380            0.18660           0.2416                0.1860          0.2750                 0.08902
2         19.69         21.25          130.00     1203.0          0.10960           0.15990          0.1974              0.12790         0.2069                 0.05999      0.7456      0.7869         4.585     94.03       0.006150        0.04006       0.03832            0.02058       0.02250                0.004571         23.57          25.53           152.50      1709.0            0.14440            0.42450           0.4504                0.2243          0.3613                 0.08758
3         11.42         20.38           77.58      386.1          0.14250           0.28390          0.2414              0.10520         0.2597                 0.09744      0.4956      1.156         3.445     27.23       0.009110        0.07458       0.05661            0.01867       0.05963                0.009208         14.91          26.50            98.87       567.7            0.20980            0.86663           0.6869                0.2575          0.6638                 0.17300
4         20.29         14.34          135.10     1297.0          0.10030           0.13280          0.1980              0.10430         0.1809                 0.05883      0.7572      0.7813         5.438     94.44       0.011490        0.02461       0.05688            0.01885       0.01756                0.005115         22.54          16.67           152.20      1575.0            0.13740            0.20500           0.4000                0.1625          0.2364                 0.07678
...       ...           ...            ...        ...              ...               ...              ...                 ...             ...                    ...        ...         ...           ...      ...            ...             ...           ...                ...           ...                    ...           ...            ...             ...         ...                ...                ...               ...                   ...             ...                    ...
564       21.56         22.39          142.00     1479.0          0.11100           0.11590          0.2439              0.13890         0.1726                 0.05623      1.044      0.8226         8.420    111.60       0.006953        0.03732       0.05494            0.01362       0.02779                0.004314         25.45          26.40           166.10      2027.0            0.14100            0.21113           0.4107                0.2216          0.2060                 0.07115
565       20.13         28.25          131.20     1261.0          0.09780           0.10340          0.1440              0.09791         0.1752                 0.05533      0.7655      2.463         5.203     99.04       0.005769        0.02423       0.04811            0.01678       0.02555                0.006033         23.69          28.610          158.70      1804.0            0.14390            0.22400           0.3226                0.1789          0.2482                 0.08772
566       16.60         28.08          108.30      858.1          0.08455           0.10023          0.0937              0.05775         0.1774                 0.05475      1.374      0.4778         9.635    105.60       0.006142        0.03492       0.04521            0.01099       0.02893                0.004837         18.98          34.12           126.70      1124.0            0.11390            0.30494           0.3744                0.1465          0.2776                 0.08019
567       20.60         29.33          140.10     1265.0          0.11780           0.27670          0.3514              0.15200         0.2397                 0.07016      0.726      1.951         5.133     99.17       0.005884        0.04302       0.05995            0.02232       0.02687                0.006643         23.69          39.28           158.00      1708.0            0.14274            0.63988           0.7267                0.2443          0.4077                 0.11213
568       7.76          24.54           47.92      181.0          0.05263           0.04362          0.0000              0.00000         0.1587                 0.05884      0.3857      1.428         2.548     19.15       0.007189        0.00000       0.00000            0.00000       0.05884                0.006351         12.02          30.37            59.60       268.6            0.09801            0.00000           0.0000                0.0000          0.1884                 0.05953
[569 rows x 30]>
```

```
df.drop(['id number', 'Unnamed: 32'], axis=1).shape
```

```
(569, 30)
```

```
df.drop(['id number', 'Unnamed: 32'], axis=1).isnull().sum()
```

```
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
dtype: object
```

```
df.drop(['id number', 'Unnamed: 32'], axis=1).corr()
```

	radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_mean	compactness_mean	concavity_mean	concave points_mean	symmetry_mean	fractal_dimension_mean	radius_se	texture_se	perimeter_se	area_se	smoothness_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
radius_mean	1.000	0.323	0.997	0.987	0.170	0.506	0.679	0.822	0.147	-0.311	0.679	0.275	0.691	0.732	0.075	0.246	0.421	0.342	0.087	-0.077	0.969	0.297	0.965	0.941	0.120	0.413	0.526	0.744	0.159	-0.062
texture_mean	0.323	1.000	0.329	0.321	-0.023	0.237	0.302	0.293	0.071	-0.076	0.275	0.386	0.281	0.295	-0.067	0.090	0.089	0.062	0.054	-0.002	0.352	0.912	0.358	0.343	0.078	0.267	0.295	0.305	0.072	0.092
perimeter_mean	0.997	0.329	1.000	0.987	0.207	0.556	0.716	0.851	0.183	-0.261	0.692	0.281	0.732	0.743	0.090	0.265	0.443	0.372	0.103	-0.057	0.970	0.303	0.970	0.941	0.150	0.455	0.563	0.771	0.189	-0.043
area_mean	0.987	0.321	0.987	1.000	0.177	0.499	0.685	0.823	0.151	-0.283	0.732	0.258	0.744	0.783	0.062	0.203	0.381	0.328	0.074	-0.092	0.959	0.287	0.959	0.958	0.123	0.390	0.512	0.722	0.143	-0.066
smoothness_mean	0.170	-0.023	0.207	0.177	1.000	0.659	0.522	0.547	0.558	0.584	0.301	0.068	0.296	0.246	0.805	0.473	0.421	0.502	0.659	0.584	0.213	0.036	0.239	0.207	0.806	0.470	0.426	0.544	0.614	0.537
compactness_mean	0.506	0.237	0.556	0.499	0.659	1.000	0.883	0.831	0.602	0.565	0.639	0.077	0.610	0.492	0.565	0.865	0.816	0.655	0.540	0.558	0.536	0.248	0.590	0.455	0.566	0.866	0.816	0.655	0.540	0.558
concavity_mean	0.679	0.302	0.716	0.685	0.522	0.883	1.000	0.921	0.500	0.336	0.688	0.308	0.715	0.658	0.448	0.754	0.884	0.661	0.438	0.390	0.687	0.299	0.729	0.640	0.448	0.754	0.884	0.661	0.438	0.390
concave points_mean	0.822	0.293	0.851	0.823	0.547	0.831	0.921	1.000	0.462	0.167	0.690	0.296	0.744	0.658	0.433	0.689	0.776	0.823	0.502	0.251	0.707	0.293	0.771	0.658	0.433	0.689	0.776	0.823	0.502	0.251
symmetry_mean	0.147	0.071	0.183	0.151	0.558	0.602	0.500	0.462	1.000	0.479	0.305	0.090	0.313	0.223	0.424	0.390	0.303	0.429	1.000	0.443	0.186	0.090	0.204	0.159	0.424	0.390	0.303	0.429	1.000	0.443
fractal_dimension_mean	-0.311	-0.076	-0.261	-0.283	0.584	0.565	0.336	0.167	0.479	1.000	-0.057	-0.046	-0.205	-0.167	0.568	0.580	0.381	0.207	0.443	1.000	-0.253	-0.051	-0.205	-0.171	0.504	0.568	0.381	0.207	0.443	1.000
radius_se	0.679	0.275	0.692	0.732	0.301	0.639	0.688	0.690	0.305	-0.057	1.000	0.213	0.972	0.951	0.166	0.497	0.632	0.568	0.058	-0.103	0.716	0.212	0.714	0.732	0.239	0.603	0.655	0.733	0.086	-0.018
texture_se	0.275	0.386	0.281	0.258	0.068	0.077	0.308	0.296	0.090	-0.046	0.213	1.000	0.215	0.186	0.101	0.268	0.295	0.267	0.047	-0.076	0.301	0.772	0.296	0.246	0.092	0.381	0.349	0.418	0.080	-0.045
perimeter_se	0.691	0.281	0.732	0.744	0.296	0.610	0.715	0.744	0.313	-0.205	0.972	0.215	1.000	0.977	0.158	0.465	0.593	0.545	0.075	-0.092	0.713	0.217	0.811	0.801	0.248	0.660	0.733	0.844	0.092	-0.028
area_se	0.732	0.295	0.743	0.783	0.246	0.492	0.658	0.658	0.223	-0.167	0.951	0.186	0.977	1.000	0.123	0.390	0.494	0.463	0.071	-0.073	0.757	0.177	0.887	0.911	0.167	0.539	0.643	0.799	0.084	-0.029
smoothness_se	0.075	-0.067	0.090	0.062	0.805	0.565	0.448	0.433	0.424	0.568	0.166	0.101	0.158	0.123	1.000	0.659	0.621	0.491	0.617	0.669	0.072	-0.046	0.092	0.055	0.746	0.501	0.549	0.445	0.620	0.707
compactness_se	0.246	0.090	0.265	0.203	0.473	0.865	0.754	0.689	0.390	0.580	0.497	0.268	0.465	0.390	0.659	1.000	0.893	0.746	0.492	0.614	0.238	0.090	0.268	0.205	0.546	0.687	0.743	0.548	0.454	0.621
concavity_se	0.421	0.089	0.443	0.381	0.421	0.816	0.884	0.776	0.303	0.381	0.632	0.295	0.593	0.494	0.621	0.893	1.000	0.771	0.384	0.492	0.268	0.092	0.386	0.287	0.543	0.746	0.943	0.771	0.423	0.568
concave points_se	0.342	0.062	0.372	0.328	0.502	0.655	0.661	0.823	0.429	0.207	0.568	0.267	0.545	0.463	0.491	0.746	0.771	1.000	0.390	0.312	0.241	0.054	0.373	0.258	0.475	0.635	0.679	0.845	0.430	0.502
symmetry_se	0.087	0.054	0.103	0.074	0.659	0.540	0.438	0.502	1.000	0.443	0.058	0.047	0.075	0.071	0.617	0.492	0.384	0.390	1.000	0.493	0.095	0.080	0.099	0.084	0.743	0.548	0.424	0.430	0.840	0.539
fractal_dimension_se	-0.077	-0.002	-0.057	-0.092	0.584	0.558	0.390	0.251	0.443	1.000	-0.103	-0.076	-0.092	-0.073	0.669	0.614	0.492	0.312	0.493	1.000	-0.128	-0.045	-0.107	-0.092	0.743	0.610	0.492	0.312	0.493	1.000
radius_worst	0.969	0.352	0.970	0.959	0.213	0.536	0.687	0.707	0.186	-0.253	0.716	0.301	0.713	0.757	0.072	0.238	0.268	0.241	0.095	-0.128	1.000	0.359	0.993	0.984	0.216	0.475	0.573	0.787	0.243	-0.080
texture_worst	0.297	0.912	0.303	0.287	0.036	0.248	0.299	0.293	0.090	-0.051	0.212	0.772	0.217	0.177	-0.046	0.090	0.092	0.054	0.080	-0.045	0.359	1.000	0.364	0.345	0.077	0.296	0.304	0.358	0.090	0.054
perimeter_worst	0.965	0.358	0.970	0.959	0.239	0.590	0.729	0.771	0.204	-0.205	0.714	0.296	0.811	0.887	0.092	0.268	0.386	0.373	0.099	-0.107	0.993	0.364	1.000	0.977	0.207	0.594	0.709	0.844	0.205	-0.092
area_worst	0.941	0.343	0.941	0.958	0.207	0.455	0.640	0.658	0.159	-0.171	0.732	0.246	0.801	0.911	0.055	0.205	0.287	0.258	0.084	-0.092	0.984	0.345	0.977	1.000	0.177	0.499	0.616	0.815	0.143	-0.092
smoothness_worst	0.120	0.078	0.150	0.123	0.806	0.566	0.448	0.433	0.424	0.504	0.239	0.092	0.248	0.167	0.746	0.546	0.543	0.475	0.743	0.743	0.216	0.077	0.207	0.177	1.000	0.568	0.518	0.547	0.623	0.743
compactness_worst	0.413	0.267	0.455	0.390	0.470	0.866	0.754	0.689	0.390	0.568	0.603	0.381	0.660	0.539	0.501	0.687	0.746	0.635	0.548	0.610	0.475	0.296	0.594	0.499	0.568	1.000	0.892	0.810	0.499	0.610
concavity_worst	0.526	0.295	0.563	0.512	0.426	0.816	0.884	0.776	0.303	0.381	0.655	0.349	0.733	0.643	0.549	0.743	0.943	0.679	0.424	0.492	0.573	0.304	0.709	0.616	0.518	0.892	1.000	0.771	0.423	0.492
concave points_worst	0.744	0.305	0.771	0.722	0.544	0.655	0.661	0.823	0.429	0.207	0.733	0.418	0.844	0.799	0.445	0.548	0.771	0.845	0.430	0.312	0.787	0.358	0.844	0.815	0.547	0.810	0.771	1.000	0.430	0.312
symmetry_worst	0.159	0.072	0.189	0.143	0.614	0.540	0.438	0.502	1.000	0.443	0.086	0.080	0.092	0.084	0.620	0.454	0.423	0.430	0.840	0.493	0.243	0.090	0.205	0.143	0.623	0.499	0.423	0.430	1.000	0.493
fractal_dimension_worst	-0.062	0.092	-0.043	-0.066	0.537	0.558	0.390	0.251	0.443	1.000	-0.018	-0.045	-0.028	-0.029	0.707	0.621	0.568	0.502	0.539	1.000	-0.080	0.054	-0.092	-0.092	0.743	0.610	0.492	0.312	0.493	1.000

```
df.drop(['id number', 'Unnamed: 32'], axis=1).corr().shape
```

```
(30, 30)
```

```
df.drop(['id number', 'Unnamed: 32'], axis=1).corr().dtypes
```

```
radius_mean                 float64
texture_mean                float64
perimeter_mean              float64
area_mean                   float64
smoothness_mean             float64
compactness_mean            float64
concavity_mean              float64
concave points_mean         float64
symmetry_mean               float64
fractal_dimension_mean      float64
radius_se                   float64
texture_se                  float64
perimeter_se                float64
area_se                     float64
smoothness_se               float64
compactness_se              float64
concavity_se                float64
concave points_se           float64
symmetry_se                 float64
fractal_dimension_se        float64
radius_worst                float64
texture_worst               float64
perimeter_worst             float64
area_worst                  float64
smoothness_worst            float64
compactness_worst           float64
concavity_worst             float64
concave points_worst        float64
symmetry_worst              float64
fractal_dimension_worst     float64
dtype: object
```

```
df.drop(['id number', 'Unnamed: 32'], axis=1).corr().isnull().sum()
```

```
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
dtype: object
```

```
Removing multicollinearity using VIF
```

```
from statsmodels.stats.outliers_influence import variance_inflation_factor
```

```
df.drop(['id number', 'Unnamed: 32'], axis=1).shape
```

```
(569, 30)
```

```
df.drop(['id number', 'Unnamed: 32'], axis=1).dtypes
```

```
radius_mean                 float64
texture_mean                float64
perimeter_mean              float64
area_mean                   float64
smoothness_mean             float64
compactness_mean            float64
concavity_mean              float64
concave points_mean         float64
symmetry_mean               float64
fractal_dimension_mean      float64
radius_se                   float64
texture_se                  float64
perimeter_se                float64
area_se                     float64
smoothness_se               float64
compactness_se              float64
concavity_se                float64
concave points_se           float64
symmetry_se                 float64
fractal_dimension_se        float64
radius_worst                float64
texture_worst               float64
perimeter_worst             float64
area_worst                  float64
smoothness_worst            float64
compactness_worst           float64
concavity_worst             float64
concave points_worst        float64
symmetry_worst              float64
fractal_dimension_worst     float64
dtype: object
```

```
df.drop(['id number', 'Unnamed: 32'], axis=1).shape[1]
```

```
30
```

```
vif = pd.DataFrame()
vif['variables'] = df.drop(['id number', 'Unnamed: 32'], axis=1).columns
vif['vif'] = [variance_inflation_factor(df.drop(['id number', 'Unnamed: 32'], axis=1).values, i)
              for i in range(df.drop(['id number', 'Unnamed: 32'], axis=1).shape[1])]
vif
```

	variables	vif
0	radius_mean	2.079441e+07
1	texture_mean	3.564259e+03
2	perimeter_mean	1.772929e+07
3	area_mean	2.177456e+03
4	smoothness_mean	1.569460e+03
5	compactness_mean	3.835968e+02
6	concavity_mean	7.929091e+02
7	concave points_mean	1.548756e+02
8	symmetry_mean	4.774223e+02
9	fractal_dimension_mean	6.188092e+02
10	radius_se	2.455032e+00
11	texture_se	3.746991e+00
12	perimeter_se	2.232062e+00
13	area_se	1.265378e+00
14	smoothness_se	1.330499e+00
15	compactness_se	1.486808e+00
16	concavity_se	1.960096e+00
17	concave points_se	1.090170e+00
18	symmetry_se	1.153451e+00
19	fractal_dimension_se	1.837648e+00
20	radius_worst	4.931891e+00
21	texture_worst	4.429233e+00
22	perimeter_worst	3.851077e+00
23	area_worst	1.395133e+00
24	smoothness_worst	1.470104e+00
25	compactness_worst	3.112080e+00
26	concavity_worst	4.294708e+00
27	concave points_worst	2.929832e+00
28	symmetry_worst	1.548480e+00
29	fractal_dimension_worst	1.805927e+00

```
df.drop(['id number', 'Unnamed: 32'], axis=1).corr().shape[1]
```

```
30
```

```
df.drop(['id number', 'Unnamed: 32'], axis=1).corr().dtypes
```

```
radius_mean                 float64
texture_mean                float64
perimeter_mean              float64
area_mean                   float64
smoothness_mean             float64
compactness_mean            float64
concavity_mean              float64
concave points_mean         float64
symmetry_mean               float64
fractal_dimension_mean      float64
radius_se                   float64
texture_se                  float64
perimeter_se                float64
area_se                     float64
smoothness_se               float64
compactness_se              float64
concavity_se                float64
concave points_se           float64
symmetry_se                 float64
fractal_dimension_se        float64
radius_worst                float64
texture_worst               float64
perimeter_worst             float64
area_worst                  float64
smoothness_worst            float64
compactness_worst           float64
concavity_worst             float64
concave points_worst        float64
symmetry_worst              float64
fractal_dimension_worst     float64
dtype: object
```

```
df.drop(['id number', 'Unnamed: 32'], axis=1).corr().isnull().sum()
```

```
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
dtype: object
```

Accuracy score using feature selection(ANOVA) and KNN is 93%

PCA dimensional reduction technique

```
from sklearn.decomposition import PCA
```

```
Scaling is compulsory
```

```
from sklearn.preprocessing import MinMaxScaler, StandardScaler
```

```
scaler_df=StandardScaler().fit_transform(df)
```

```
scaler_df
```

```
array([[ 1.09706398, -2.07333501,  1.26993369,  0.984375  ,  1.56846633,
         3.28351467,  2.65287398,  2.15072364,  0.39053252,  -0.04823965,
         1.09448818, -0.48877065,  1.28062512,  0.98043054,  0.35987117,
         1.91421825,  1.405506  ,  1.00384756,  1.315929  ,  0.51160537,
         0.23328963, -1.83097127,  1.62928448,  1.37654431,  0.14064627,
         3.37119392,  3.20724431,  2.29607613,  2.748692  ,  1.935992  ,
         2.175786  ,  0.62331909],
       [ 1.829821  , -0.353632  ,  1.685955  ,  1.908708  , -0.826962  ,
        -0.487072  , -0.023846  ,  0.548144  ,  0.001392  , -0.868652  ,
         2.735696  , -0.746811  ,  2.777042  ,  2.289577  , -0.745894  ,
         -0.048205  , -0.309380  , -0.744232  , -0.202240  , -1.090823  ,
         1.084405  , -1.359293  ,  1.909249  ,  1.636962  , -0.854911  ,
         -0.826962  , -0.487072  , -0.023846  ,  0.548144  ,  0.001392  ,
         -0.868652  , -0.307796  ],
       [ 1.579888  , -0.456187  ,  1.565954  ,  1.557513  ,  0.942210  ,
         1.052926  ,  1.363478  ,  2.035975  ,  0.939685  ,  -0.398008  ,
         1.915032  , -0.048189  ,  1.931759  ,  1.909083  ,  0.377162  ,
         1.058611  ,  1.447951  ,  1.290320  ,  0.745760  ,  0.772608  ,
         1.379401  , -0.232790  ,  1.475766  ,  1.431403  ,  0.965626  ,
         1.052926  ,  1.363478  ,  2.035975  ,  0.939685  , -0.398008  ,
         1.915032  , -0.048189  ],
       [ 0.70228   , -1.828227  ,  0.693043  ,  0.596384  ,  1.837308  ,
         1.466770  ,  0.854274  ,  1.405161  ,  0.267232  ,  1.058611  ,
         0.067351  , -1.090064  ,  0.067351  ,  0.067351  ,  1.637893  ,
         0.335633  ,  0.331393  ,  0.358929  ,  0.023274  ,  0.728026  ,
         0.135777  , -1.057071  ,  0.135777  ,  0.135777  ,  2.004957  ,
         1.466770  ,  0.854274  ,  1.405161  ,  0.267232  ,  1.058611  ,
         0.067351  , -1.090064  ],
       [ 1.307686  , -0.768909  ,  1.296095  ,  1.268817  ,  0.641069  ,
         0.779655  ,  1.274431  ,  1.620776  ,  0.350714  ,  0.283556  ,
         0.201259  , -0.048255  ,  0.202706  ,  0.167039  ,  0.506501  ,
         0.212575  ,  0.376223  ,  0.470602  ,  0.274031  ,  0.062
```