Project-2: Breast_Cancer_Prediction

Importing the libraries

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns


from sklearn.datasets import load_breast_cancer
```

Loading the Dataset

```
# Loading the Breast Cancer data in a variable
cancer = load_breast_cancer()

# printing the data in a dictionary format
cancer
```

```
{'data': array([[1.799e+01, 1.038e+01, 1.228e+02, ..., 2.654e-01,
4.601e-01,
         1.189e-01],
        [2.057e+01, 1.777e+01, 1.329e+02, ..., 1.860e-01, 2.750e-01,
         8.902e-02],
        [1.969e+01, 2.125e+01, 1.300e+02, ..., 2.430e-01, 3.613e-01,
         8.758e-02],
        ...,
        [1.660e+01, 2.808e+01, 1.083e+02, ..., 1.418e-01, 2.218e-01,
         7.820e-02],
        [2.060e+01, 2.933e+01, 1.401e+02, ..., 2.650e-01, 4.087e-01,
         1.240e-01],
        [7.760e+00, 2.454e+01, 4.792e+01, ..., 0.000e+00, 2.871e-01,
         7.039e-02]]),
 'target': array([0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 1, 1, 1,
        0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0,
0,
        0, 0, 1, 0, 1, 1, 1, 1, 1, 0, 0, 1, 0, 0, 1, 1, 1, 1, 0, 1, 0,
0,
        1, 1, 1, 1, 0, 1, 0, 0, 1, 0, 1, 0, 0, 1, 1, 1, 0, 0, 1, 0, 0,
0,
        1, 1, 1, 0, 1, 1, 0, 0, 1, 1, 1, 0, 0, 1, 1, 1, 1, 0, 1, 1, 0,
1,
        1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 1, 0, 0, 1, 1, 1, 0, 0, 1, 0, 1,
0,
        0, 1, 0, 0, 1, 1, 0, 1, 1, 0, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1,
1,
        1, 1, 0, 1, 1, 1, 1, 0, 0, 1, 0, 1, 1, 0, 0, 1, 1, 0, 0, 1, 1,
```

```
1,
        1, 0, 1, 1, 0, 0, 0, 1, 0, 1, 0, 1, 1, 1, 0, 1, 1, 0, 0, 1, 0,
0,
        0, 0, 1, 0, 0, 0, 1, 0, 1, 0, 1, 1, 0, 1, 0, 0, 0, 0, 1, 1, 0,
0,
        1, 1, 1, 0, 1, 1, 1, 1, 1, 0, 0, 1, 1, 0, 1, 1, 0, 0, 1, 0, 1,
1,
        1, 1, 0, 1, 1, 1, 1, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0,
        0, 0, 1, 1, 1, 1, 1, 1, 0, 1, 0, 1, 1, 0, 1, 1, 0, 1, 0, 0, 1,
1,
        1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 0, 1, 0, 1, 1, 1, 1,
1,
        1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 0, 1, 0, 1, 1, 1, 1, 0,
0,
        0, 1, 1, 1, 1, 0, 1, 0, 1, 0, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1,
0,
        0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 1, 0, 0, 0, 1, 0,
0,
        1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 0, 1, 1, 1, 0, 1, 1, 0, 0, 1,
1,
        1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 0, 1, 1,
0,
        1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 0, 0, 1, 0, 1, 1, 1,
1,
        1, 0, 1, 1, 0, 1, 0, 1, 1, 0, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 0,
0,
        1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1,
1,
        1, 1, 1, 0, 1, 0, 1, 1, 0, 1, 1, 1, 1, 1, 0, 0, 1, 0, 1, 0, 1,
1,
        1, 1, 1, 0, 1, 1, 0, 1, 0, 1, 0, 0, 1, 1, 1, 0, 1, 1, 1, 1, 1,
1,
        1, 1, 1, 1, 1, 0, 1, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
1,
        1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 1]),
 'frame': None,
 'target_names': array(['malignant', 'benign'], dtype='<U9'),
 'DESCR': '.. _breast_cancer_dataset:\n\nBreast cancer wisconsin
(diagnostic) dataset\n--------------------------------------------\n\
n**Data Set Characteristics:**\n\n    :Number of Instances: 569\n\n
:Number of Attributes: 30 numeric, predictive attributes and the
class\n\n    :Attribute Information:\n        - radius (mean of
distances from center to points on the perimeter)\n        - texture
(standard deviation of gray-scale values)\n        - perimeter\n
- area\n        - smoothness (local variation in radius lengths)\n
- compactness (perimeter^2 / area - 1.0)\n        - concavity
(severity of concave portions of the contour)\n        - concave
points (number of concave portions of the contour)\n        -
```

```
symmetry\n        - fractal dimension ("coastline approximation" - 1)\
n\n        The mean, standard error, and "worst" or largest (mean of
the three\n        worst/largest values) of these features were
computed for each image,\n        resulting in 30 features.  For
instance, field 0 is Mean Radius, field\n        10 is Radius SE,
field 20 is Worst Radius.\n\n        - class:\n                - WDBC-
Malignant\n                - WDBC-Benign\n\n    :Summary Statistics:\
n\n    ===================================== ====== ======\n
Min    Max\n    ===================================== ====== ======\n
radius (mean):                        6.981  28.11\n    texture
(mean):                       9.71   39.28\n    perimeter (mean):
43.79  188.5\n    area (mean):                         143.5  2501.0\
n    smoothness (mean):                    0.053  0.163\n
compactness (mean):                   0.019  0.345\n    concavity
(mean):                     0.0    0.427\n    concave points (mean):
0.0    0.201\n    symmetry (mean):                     0.106  0.304\n
fractal dimension (mean):             0.05   0.097\n    radius
(standard error):              0.112  2.873\n    texture (standard
error):            0.36   4.885\n    perimeter (standard error):
0.757  21.98\n    area (standard error):               6.802  542.2\n
smoothness (standard error):          0.002  0.031\n    compactness
(standard error):         0.002  0.135\n    concavity (standard
error):           0.0    0.396\n    concave points (standard error):
0.0    0.053\n    symmetry (standard error):           0.008  0.079\n
fractal dimension (standard error):   0.001  0.03\n    radius (worst):
7.93   36.04\n    texture (worst):                     12.02  49.54\n
perimeter (worst):                    50.41  251.2\n    area (worst):
185.2  4254.0\n    smoothness (worst):                  0.071  0.223\
n    compactness (worst):                  0.027  1.058\n    concavity
(worst):                    0.0    1.252\n    concave points (worst):
0.0    0.291\n    symmetry (worst):                    0.156  0.664\n
fractal dimension (worst):            0.055  0.208\n
===================================== ====== ======\n\n    :Missing
Attribute Values: None\n\n    :Class Distribution: 212 - Malignant,
357 - Benign\n\n    :Creator:  Dr. William H. Wolberg, W. Nick Street,
Olvi L. Mangasarian\n\n    :Donor: Nick Street\n\n    :Date: November,
1995\n\nThis is a copy of UCI ML Breast Cancer Wisconsin (Diagnostic)
datasets.\nhttps://goo.gl/U2Uwz2\n\nFeatures are computed from a
digitized image of a fine needle\naspirate (FNA) of a breast mass.
They describe\ncharacteristics of the cell nuclei present in the
image.\n\nSeparating plane described above was obtained using\
nMultisurface Method-Tree (MSM-T) [K. P. Bennett, "Decision Tree\
nConstruction Via Linear Programming." Proceedings of the 4th\nMidwest
Artificial Intelligence and Cognitive Science Society,\npp. 97-101,
1992], a classification method which uses linear\nprogramming to
construct a decision tree.  Relevant features\nwere selected using an
exhaustive search in the space of 1-4\nfeatures and 1-3 separating
planes.\n\nThe actual linear program used to obtain the separating
plane\nin the 3-dimensional space is that described in:\n[K. P.
```

```
Bennett and O. L. Mangasarian: "Robust Linear\nProgramming
Discrimination of Two Linearly Inseparable Sets",\nOptimization
Methods and Software 1, 1992, 23-34].\n\nThis database is also
available through the UW CS ftp server:\n\nftp ftp.cs.wisc.edu\ncd
math-prog/cpo-dataset/machine-learn/WDBC/\n\n.. topic:: References\n\n
- W.N. Street, W.H. Wolberg and O.L. Mangasarian. Nuclear feature
extraction \n     for breast tumor diagnosis. IS&T/SPIE 1993
International Symposium on \n     Electronic Imaging: Science and
Technology, volume 1905, pages 861-870,\n     San Jose, CA, 1993.\n
- O.L. Mangasarian, W.N. Street and W.H. Wolberg. Breast cancer
diagnosis and \n     prognosis via linear programming. Operations
Research, 43(4), pages 570-577, \n     July-August 1995.\n   - W.H.
Wolberg, W.N. Street, and O.L. Mangasarian. Machine learning
techniques\n     to diagnose breast cancer from fine-needle aspirates.
Cancer Letters 77 (1994) \n     163-171.',
 'feature_names': array(['mean radius', 'mean texture', 'mean
perimeter', 'mean area',
        'mean smoothness', 'mean compactness', 'mean concavity',
        'mean concave points', 'mean symmetry', 'mean fractal
dimension',
        'radius error', 'texture error', 'perimeter error', 'area
error',
        'smoothness error', 'compactness error', 'concavity error',
        'concave points error', 'symmetry error',
        'fractal dimension error', 'worst radius', 'worst texture',
        'worst perimeter', 'worst area', 'worst smoothness',
        'worst compactness', 'worst concavity', 'worst concave
points',
        'worst symmetry', 'worst fractal dimension'], dtype='<U23'),
 'filename': 'breast_cancer.csv',
 'data_module': 'sklearn.datasets.data'}
```

```
# printing the keys of the dictionary to get enough details about the
data
cancer.keys()
```

```
dict_keys(['data', 'target', 'frame', 'target_names', 'DESCR',
'feature_names', 'filename', 'data_module'])
```

```
cancer.values()
```

```
dict_values([array([[1.799e+01, 1.038e+01, 1.228e+02, ..., 2.654e-01,
4.601e-01,
        1.189e-01],
       [2.057e+01, 1.777e+01, 1.329e+02, ..., 1.860e-01, 2.750e-01,
        8.902e-02],
       [1.969e+01, 2.125e+01, 1.300e+02, ..., 2.430e-01, 3.613e-01,
        8.758e-02],
       ...,
       [1.660e+01, 2.808e+01, 1.083e+02, ..., 1.418e-01, 2.218e-01,
```

```
         7.820e-02],
        [2.060e+01, 2.933e+01, 1.401e+02, ..., 2.650e-01, 4.087e-01,
         1.240e-01],
        [7.760e+00, 2.454e+01, 4.792e+01, ..., 0.000e+00, 2.871e-01,
         7.039e-02]]), array([0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 1, 1, 1,
        0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0,
0,
        0, 0, 1, 0, 1, 1, 1, 1, 1, 0, 0, 1, 0, 0, 1, 1, 1, 1, 0, 1, 0,
0,
        1, 1, 1, 1, 0, 1, 0, 0, 1, 0, 1, 0, 0, 1, 1, 1, 0, 0, 1, 0, 0,
0,
        1, 1, 1, 0, 1, 1, 0, 0, 1, 1, 1, 0, 0, 1, 1, 1, 1, 0, 1, 1, 0,
1,
        1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 1, 0, 0, 1, 1, 1, 0, 0, 1, 0, 1,
0,
        0, 1, 0, 0, 1, 1, 0, 1, 1, 0, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1,
1,
        1, 1, 0, 1, 1, 1, 1, 0, 0, 1, 0, 1, 1, 0, 0, 1, 1, 0, 0, 1, 1,
1,
        1, 0, 1, 1, 0, 0, 0, 1, 0, 1, 0, 1, 1, 1, 0, 1, 1, 0, 0, 1, 0,
0,
        0, 0, 1, 0, 0, 0, 1, 0, 1, 0, 1, 1, 0, 1, 0, 0, 0, 0, 1, 1, 0,
0,
        1, 1, 1, 0, 1, 1, 1, 1, 1, 0, 0, 1, 1, 0, 1, 1, 0, 0, 1, 0, 1,
1,
        1, 1, 0, 1, 1, 1, 1, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0,
        0, 0, 1, 1, 1, 1, 1, 1, 0, 1, 0, 1, 1, 0, 1, 1, 0, 1, 0, 0, 1,
1,
        1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 0, 1, 0, 1, 1, 1, 1,
1,
        1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 0, 1, 0, 1, 1, 1, 1, 0,
0,
        0, 1, 1, 1, 1, 0, 1, 0, 1, 0, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1,
0,
        0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 1, 0, 0, 0, 1, 0,
0,
        1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 0, 1, 1, 1, 0, 1, 1, 0, 0, 1,
1,
        1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 0, 1, 1,
0,
        1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 0, 0, 1, 0, 1, 1, 1,
1,
        1, 0, 1, 1, 0, 1, 0, 1, 1, 0, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 0,
0,
        1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1,
1,
        1, 1, 1, 0, 1, 0, 1, 1, 0, 1, 1, 1, 1, 1, 0, 0, 1, 0, 1, 0, 1,
```

```
1,
       1, 1, 1, 0, 1, 1, 0, 1, 0, 1, 0, 0, 1, 1, 1, 0, 1, 1, 1, 1, 1,
1,
       1, 1, 1, 1, 1, 0, 1, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
1,
       1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 1]),
None, array(['malignant', 'benign'], dtype='<U9'), '..
_breast_cancer_dataset:\n\nBreast cancer wisconsin (diagnostic)
dataset\n--------------------------------------------\n\n**Data Set
Characteristics:**\n\n    :Number of Instances: 569\n\n    :Number of
Attributes: 30 numeric, predictive attributes and the class\n\
n    :Attribute Information:\n        - radius (mean of distances from
center to points on the perimeter)\n        - texture (standard
deviation of gray-scale values)\n        - perimeter\n        - area\n
- smoothness (local variation in radius lengths)\n        -
compactness (perimeter^2 / area - 1.0)\n        - concavity (severity
of concave portions of the contour)\n        - concave points (number
of concave portions of the contour)\n        - symmetry\n        -
fractal dimension ("coastline approximation" - 1)\n\n        The mean,
standard error, and "worst" or largest (mean of the three\n
worst/largest values) of these features were computed for each image,\
n        resulting in 30 features.  For instance, field 0 is Mean
Radius, field\n        10 is Radius SE, field 20 is Worst Radius.\n\n
- class:\n                - WDBC-Malignant\n                - WDBC-
Benign\n\n    :Summary Statistics:\n\n
===================================== ====== ======\n
Min    Max\n    ===================================== ====== ======\n
radius (mean):                        6.981  28.11\n    texture
(mean):                       9.71   39.28\n    perimeter (mean):
43.79  188.5\n    area (mean):                        143.5  2501.0\
n    smoothness (mean):                    0.053  0.163\n
compactness (mean):                   0.019  0.345\n    concavity
(mean):                     0.0    0.427\n    concave points (mean):
0.0    0.201\n    symmetry (mean):                      0.106  0.304\n
fractal dimension (mean):             0.05   0.097\n    radius
(standard error):             0.112  2.873\n    texture (standard
error):              0.36   4.885\n    perimeter (standard error):
0.757  21.98\n    area (standard error):                6.802  542.2\n
smoothness (standard error):          0.002  0.031\n    compactness
(standard error):         0.002  0.135\n    concavity (standard
error):            0.0    0.396\n    concave points (standard error):
0.0    0.053\n    symmetry (standard error):            0.008  0.079\n
fractal dimension (standard error):   0.001  0.03\n    radius (worst):
7.93   36.04\n    texture (worst):                      12.02  49.54\n
perimeter (worst):                    50.41  251.2\n    area (worst):
185.2  4254.0\n    smoothness (worst):                   0.071  0.223\
n    compactness (worst):                  0.027  1.058\n    concavity
(worst):                    0.0    1.252\n    concave points (worst):
0.0    0.291\n    symmetry (worst):                      0.156  0.664\n
```

```
fractal dimension (worst):            0.055  0.208\n
===================================== ====== ======\n\n    :Missing
Attribute Values: None\n\n    :Class Distribution: 212 - Malignant,
357 - Benign\n\n    :Creator:  Dr. William H. Wolberg, W. Nick Street,
Olvi L. Mangasarian\n\n    :Donor: Nick Street\n\n    :Date: November,
1995\n\nThis is a copy of UCI ML Breast Cancer Wisconsin (Diagnostic)
datasets.\nhttps://goo.gl/U2Uwz2\n\nFeatures are computed from a
digitized image of a fine needle\naspirate (FNA) of a breast mass.
They describe\ncharacteristics of the cell nuclei present in the
image.\n\nSeparating plane described above was obtained using\
nMultisurface Method-Tree (MSM-T) [K. P. Bennett, "Decision Tree\
nConstruction Via Linear Programming." Proceedings of the 4th\nMidwest
Artificial Intelligence and Cognitive Science Society,\npp. 97-101,
1992], a classification method which uses linear\nprogramming to
construct a decision tree.  Relevant features\nwere selected using an
exhaustive search in the space of 1-4\nfeatures and 1-3 separating
planes.\n\nThe actual linear program used to obtain the separating
plane\nin the 3-dimensional space is that described in:\n[K. P.
Bennett and O. L. Mangasarian: "Robust Linear\nProgramming
Discrimination of Two Linearly Inseparable Sets",\nOptimization
Methods and Software 1, 1992, 23-34].\n\nThis database is also
available through the UW CS ftp server:\n\nftp ftp.cs.wisc.edu\ncd
math-prog/cpo-dataset/machine-learn/WDBC/\n\n.. topic:: References\n\n
- W.N. Street, W.H. Wolberg and O.L. Mangasarian. Nuclear feature
extraction \n     for breast tumor diagnosis. IS&T/SPIE 1993
International Symposium on \n     Electronic Imaging: Science and
Technology, volume 1905, pages 861-870,\n     San Jose, CA, 1993.\n
- O.L. Mangasarian, W.N. Street and W.H. Wolberg. Breast cancer
diagnosis and \n     prognosis via linear programming. Operations
Research, 43(4), pages 570-577, \n     July-August 1995.\n   - W.H.
Wolberg, W.N. Street, and O.L. Mangasarian. Machine learning
techniques\n     to diagnose breast cancer from fine-needle aspirates.
Cancer Letters 77 (1994) \n     163-171.', array(['mean radius', 'mean
texture', 'mean perimeter', 'mean area',
       'mean smoothness', 'mean compactness', 'mean concavity',
       'mean concave points', 'mean symmetry', 'mean fractal
dimension',
       'radius error', 'texture error', 'perimeter error', 'area
error',
       'smoothness error', 'compactness error', 'concavity error',
       'concave points error', 'symmetry error',
       'fractal dimension error', 'worst radius', 'worst texture',
       'worst perimeter', 'worst area', 'worst smoothness',
       'worst compactness', 'worst concavity', 'worst concave points',
       'worst symmetry', 'worst fractal dimension'], dtype='<U23'),
'breast_cancer.csv', 'sklearn.datasets.data'])

# description of breast cancer dataset
cancer["DESCR"]
```

```
'.. _breast_cancer_dataset:\n\nBreast cancer wisconsin (diagnostic)
dataset\n--------------------------------------------\n\n**Data Set
Characteristics:**\n\n    :Number of Instances: 569\n\n    :Number of
Attributes: 30 numeric, predictive attributes and the class\n\
n    :Attribute Information:\n        - radius (mean of distances from
center to points on the perimeter)\n        - texture (standard
deviation of gray-scale values)\n        - perimeter\n        - area\n
- smoothness (local variation in radius lengths)\n        -
compactness (perimeter^2 / area - 1.0)\n        - concavity (severity
of concave portions of the contour)\n        - concave points (number
of concave portions of the contour)\n        - symmetry\n        -
fractal dimension ("coastline approximation" - 1)\n\n        The mean,
standard error, and "worst" or largest (mean of the three\n
worst/largest values) of these features were computed for each image,\
n        resulting in 30 features.  For instance, field 0 is Mean
Radius, field\n        10 is Radius SE, field 20 is Worst Radius.\n\n
- class:\n                - WDBC-Malignant\n                - WDBC-
Benign\n\n    :Summary Statistics:\n\n
===================================== ====== ======\n
Min    Max\n    ===================================== ====== ======\n
radius (mean):                        6.981  28.11\n    texture
(mean):                        9.71   39.28\n    perimeter (mean):
43.79  188.5\n    area (mean):                        143.5  2501.0\
n    smoothness (mean):                    0.053  0.163\n
compactness (mean):                   0.019  0.345\n    concavity
(mean):                      0.0    0.427\n    concave points (mean):
0.0    0.201\n    symmetry (mean):                      0.106  0.304\n
fractal dimension (mean):             0.05   0.097\n    radius
(standard error):              0.112  2.873\n    texture (standard
error):           0.36   4.885\n    perimeter (standard error):
0.757  21.98\n    area (standard error):                6.802  542.2\n
smoothness (standard error):          0.002  0.031\n    compactness
(standard error):         0.002  0.135\n    concavity (standard
error):          0.0    0.396\n    concave points (standard error):
0.0    0.053\n    symmetry (standard error):            0.008  0.079\n
fractal dimension (standard error):   0.001  0.03\n    radius (worst):
7.93   36.04\n    texture (worst):                      12.02  49.54\n
perimeter (worst):                    50.41  251.2\n    area (worst):
185.2  4254.0\n    smoothness (worst):                   0.071  0.223\
n    compactness (worst):                  0.027  1.058\n    concavity
(worst):                     0.0    1.252\n    concave points (worst):
0.0    0.291\n    symmetry (worst):                     0.156  0.664\n
fractal dimension (worst):            0.055  0.208\n
===================================== ====== ======\n\n    :Missing
Attribute Values: None\n\n    :Class Distribution: 212 - Malignant,
357 - Benign\n\n    :Creator:  Dr. William H. Wolberg, W. Nick Street,
Olvi L. Mangasarian\n\n    :Donor: Nick Street\n\n    :Date: November,
1995\n\nThis is a copy of UCI ML Breast Cancer Wisconsin (Diagnostic)
datasets.\nhttps://goo.gl/U2Uwz2\n\nFeatures are computed from a
digitized image of a fine needle\naspirate (FNA) of a breast mass.
```

```
They describe\ncharacteristics of the cell nuclei present in the
image.\n\nSeparating plane described above was obtained using\
nMultisurface Method-Tree (MSM-T) [K. P. Bennett, "Decision Tree\
nConstruction Via Linear Programming." Proceedings of the 4th\nMidwest
Artificial Intelligence and Cognitive Science Society,\npp. 97-101,
1992], a classification method which uses linear\nprogramming to
construct a decision tree.  Relevant features\nwere selected using an
exhaustive search in the space of 1-4\nfeatures and 1-3 separating
planes.\n\nThe actual linear program used to obtain the separating
plane\nin the 3-dimensional space is that described in:\n[K. P.
Bennett and O. L. Mangasarian: "Robust Linear\nProgramming
Discrimination of Two Linearly Inseparable Sets",\nOptimization
Methods and Software 1, 1992, 23-34].\n\nThis database is also
available through the UW CS ftp server:\n\nftp ftp.cs.wisc.edu\ncd
math-prog/cpo-dataset/machine-learn/WDBC/\n\n.. topic:: References\n\n
- W.N. Street, W.H. Wolberg and O.L. Mangasarian. Nuclear feature
extraction \n     for breast tumor diagnosis. IS&T/SPIE 1993
International Symposium on \n     Electronic Imaging: Science and
Technology, volume 1905, pages 861-870,\n     San Jose, CA, 1993.\n
- O.L. Mangasarian, W.N. Street and W.H. Wolberg. Breast cancer
diagnosis and \n     prognosis via linear programming. Operations
Research, 43(4), pages 570-577, \n     July-August 1995.\n   - W.H.
Wolberg, W.N. Street, and O.L. Mangasarian. Machine learning
techniques\n     to diagnose breast cancer from fine-needle aspirates.
Cancer Letters 77 (1994) \n     163-171.'
```

```python
# printing the target column which is either 0=Malignenet or 1=Benign
print(cancer['target'])
```

```
[0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0
 1 0 0 0 0 0 0 0 0 1 0 1 1 1 1 1 0 0 1 0 0 1 1 1 1 0 1 0 0 1 1 1 1 0 1
0 0
 1 0 1 0 0 1 1 1 0 0 1 0 0 0 1 1 1 0 1 1 0 0 1 1 1 0 0 1 1 1 1 0 1 1 0
1 1
 1 1 1 1 1 1 0 0 0 1 0 0 1 1 1 0 0 1 0 1 0 0 1 0 0 1 1 0 1 1 0 1 1 1 1
0 1
 1 1 1 1 1 1 1 1 0 1 1 1 1 0 0 1 0 1 1 0 0 1 1 0 0 1 1 1 1 0 1 1 0 0 0
1 0
 1 0 1 1 1 0 1 1 0 0 1 0 0 0 0 1 0 0 0 1 0 1 0 1 1 0 1 0 0 0 0 1 1 0 0
1 1
 1 0 1 1 1 1 1 0 0 1 1 0 1 1 0 0 1 0 1 1 1 1 0 1 1 1 1 1 0 1 0 0 0 0 0
0 0
 0 0 0 0 0 0 0 1 1 1 1 1 1 0 1 0 1 1 0 1 1 0 1 0 0 1 1 1 1 1 1 1 1 1 1
1 1
 1 0 1 1 0 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 0 1 0 1 1 1 1 0 0 0
1 1
 1 1 0 1 0 1 0 1 1 1 0 1 1 1 1 1 1 1 0 0 0 1 1 1 1 1 1 1 1 1 1 1 0 0 1
0 0
 0 1 0 0 1 1 1 1 1 0 1 1 1 1 1 0 1 1 1 0 1 1 0 0 1 1 1 1 1 1 0 1 1 1 1
```

```
1 1
 1 0 1 1 1 1 1 0 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 0 1 0 0 1 0 1 1 1 1 1 0
1 1
 0 1 0 1 1 0 1 0 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1
0 1
 1 1 1 1 1 1 0 1 0 1 1 0 1 1 1 1 1 0 0 1 0 1 0 1 1 1 1 1 0 1 1 0 1 0 1
0 0
 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 0 1 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
1 1
 1 1 1 1 1 1 1 0 0 0 0 0 0 1]

# printing the target names
print(cancer['target_names'])

['malignant' 'benign']

# printing the alll the columns name which are also know as features
print(cancer['feature_names'])

['mean radius' 'mean texture' 'mean perimeter' 'mean area'
 'mean smoothness' 'mean compactness' 'mean concavity'
 'mean concave points' 'mean symmetry' 'mean fractal dimension'
 'radius error' 'texture error' 'perimeter error' 'area error'
 'smoothness error' 'compactness error' 'concavity error'
 'concave points error' 'symmetry error' 'fractal dimension error'
 'worst radius' 'worst texture' 'worst perimeter' 'worst area'
 'worst smoothness' 'worst compactness' 'worst concavity'
 'worst concave points' 'worst symmetry' 'worst fractal dimension']

df_cancer = pd.DataFrame(np.c_[cancer['data'], cancer['target']],
columns= np.append(cancer['feature_names'],['target']))

df_cancer

     mean radius  mean texture  mean perimeter  mean area  mean
smoothness  \
0          17.99         10.38          122.80     1001.0
0.11840
1          20.57         17.77          132.90     1326.0
0.08474
2          19.69         21.25          130.00     1203.0
0.10960
3          11.42         20.38           77.58      386.1
0.14250
4          20.29         14.34          135.10     1297.0
0.10030
..           ...           ...             ...        ...
...
564        21.56         22.39          142.00     1479.0
0.11100
565        20.13         28.25          131.20     1261.0
```

```
0.09780
566        16.60         28.08          108.30      858.1
0.08455
567        20.60         29.33          140.10     1265.0
0.11780
568         7.76         24.54           47.92      181.0
0.05263

     mean compactness  mean concavity  mean concave points  mean
symmetry  \
0             0.27760         0.30010              0.14710
0.2419
1             0.07864         0.08690              0.07017
0.1812
2             0.15990         0.19740              0.12790
0.2069
3             0.28390         0.24140              0.10520
0.2597
4             0.13280         0.19800              0.10430
0.1809
..                ...             ...                  ...
...
564           0.11590         0.24390              0.13890
0.1726
565           0.10340         0.14400              0.09791
0.1752
566           0.10230         0.09251              0.05302
0.1590
567           0.27700         0.35140              0.15200
0.2397
568           0.04362         0.00000              0.00000
0.1587

     mean fractal dimension  ...  worst texture  worst perimeter
worst area  \
0                   0.07871  ...          17.33           184.60
2019.0
1                   0.05667  ...          23.41           158.80
1956.0
2                   0.05999  ...          25.53           152.50
1709.0
3                   0.09744  ...          26.50            98.87
567.7
4                   0.05883  ...          16.67           152.20
1575.0
..                      ...  ...            ...              ...
...
564                 0.05623  ...          26.40           166.10
2027.0
```

```
565                  0.05533  ...          38.25           155.00
1731.0
566                  0.05648  ...          34.12           126.70
1124.0
567                  0.07016  ...          39.42           184.60
1821.0
568                  0.05884  ...          30.37            59.16
268.6

     worst smoothness  worst compactness  worst concavity  \
0             0.16220            0.66560           0.7119
1             0.12380            0.18660           0.2416
2             0.14440            0.42450           0.4504
3             0.20980            0.86630           0.6869
4             0.13740            0.20500           0.4000
..                ...                ...              ...
564           0.14100            0.21130           0.4107
565           0.11660            0.19220           0.3215
566           0.11390            0.30940           0.3403
567           0.16500            0.86810           0.9387
568           0.08996            0.06444           0.0000

     worst concave points  worst symmetry  worst fractal dimension
target
0                  0.2654          0.4601                  0.11890
0.0
1                  0.1860          0.2750                  0.08902
0.0
2                  0.2430          0.3613                  0.08758
0.0
3                  0.2575          0.6638                  0.17300
0.0
4                  0.1625          0.2364                  0.07678
0.0
..                    ...             ...                      ...
...
564                0.2216          0.2060                  0.07115
0.0
565                0.1628          0.2572                  0.06637
0.0
566                0.1418          0.2218                  0.07820
0.0
567                0.2650          0.4087                  0.12400
0.0
568                0.0000          0.2871                  0.07039
1.0

[569 rows x 31 columns]
```

Performing Data Preprocessing

```
df_cancer.head()

   mean radius  mean texture  mean perimeter  mean area  mean
smoothness  \
0        17.99         10.38          122.80     1001.0
0.11840
1        20.57         17.77          132.90     1326.0
0.08474
2        19.69         21.25          130.00     1203.0
0.10960
3        11.42         20.38           77.58      386.1
0.14250
4        20.29         14.34          135.10     1297.0
0.10030

   mean compactness  mean concavity  mean concave points  mean
symmetry  \
0           0.27760          0.3001              0.14710
0.2419
1           0.07864          0.0869              0.07017
0.1812
2           0.15990          0.1974              0.12790
0.2069
3           0.28390          0.2414              0.10520
0.2597
4           0.13280          0.1980              0.10430
0.1809

   mean fractal dimension  ...  worst texture  worst perimeter  worst
area  \
0                 0.07871  ...          17.33           184.60
2019.0
1                 0.05667  ...          23.41           158.80
1956.0
2                 0.05999  ...          25.53           152.50
1709.0
3                 0.09744  ...          26.50            98.87
567.7
4                 0.05883  ...          16.67           152.20
1575.0

   worst smoothness  worst compactness  worst concavity  worst concave
points  \
0            0.1622             0.6656           0.7119
0.2654
1            0.1238             0.1866           0.2416
0.1860
2            0.1444             0.4245           0.4504
```

```
0.2430
3           0.2098             0.8663           0.6869
0.2575
4           0.1374             0.2050           0.4000
0.1625

   worst symmetry  worst fractal dimension  target
0          0.4601                  0.11890     0.0
1          0.2750                  0.08902     0.0
2          0.3613                  0.08758     0.0
3          0.6638                  0.17300     0.0
4          0.2364                  0.07678     0.0

[5 rows x 31 columns]
```

```
df_cancer.tail()
```

```
     mean radius  mean texture  mean perimeter  mean area  mean
smoothness  \
564        21.56         22.39          142.00     1479.0
0.11100
565        20.13         28.25          131.20     1261.0
0.09780
566        16.60         28.08          108.30      858.1
0.08455
567        20.60         29.33          140.10     1265.0
0.11780
568         7.76         24.54           47.92      181.0
0.05263

     mean compactness  mean concavity  mean concave points  mean
symmetry  \
564           0.11590         0.24390              0.13890
0.1726
565           0.10340         0.14400              0.09791
0.1752
566           0.10230         0.09251              0.05302
0.1590
567           0.27700         0.35140              0.15200
0.2397
568           0.04362         0.00000              0.00000
0.1587

     mean fractal dimension  ...  worst texture  worst perimeter
worst area  \
564                 0.05623  ...          26.40           166.10
2027.0
565                 0.05533  ...          38.25           155.00
1731.0
566                 0.05648  ...          34.12           126.70
```

```
1124.0
567                  0.07016  ...          39.42          184.60
1821.0
568                  0.05884  ...          30.37           59.16
268.6

     worst smoothness  worst compactness  worst concavity  \
564           0.14100            0.21130           0.4107
565           0.11660            0.19220           0.3215
566           0.11390            0.30940           0.3403
567           0.16500            0.86810           0.9387
568           0.08996            0.06444           0.0000

     worst concave points  worst symmetry  worst fractal dimension
target
564                0.2216          0.2060                  0.07115
0.0
565                0.1628          0.2572                  0.06637
0.0
566                0.1418          0.2218                  0.07820
0.0
567                0.2650          0.4087                  0.12400
0.0
568                0.0000          0.2871                  0.07039
1.0

[5 rows x 31 columns]
```

```
df_cancer.describe()
```

```
       mean radius  mean texture  mean perimeter    mean area  \
count   569.000000    569.000000      569.000000   569.000000
mean     14.127292     19.289649       91.969033   654.889104
std       3.524049      4.301036       24.298981   351.914129
min       6.981000      9.710000       43.790000   143.500000
25%      11.700000     16.170000       75.170000   420.300000
50%      13.370000     18.840000       86.240000   551.100000
75%      15.780000     21.800000      104.100000   782.700000
max      28.110000     39.280000      188.500000  2501.000000

       mean smoothness  mean compactness  mean concavity  mean concave
points  \
count       569.000000        569.000000      569.000000
569.000000
mean          0.096360          0.104341        0.088799
0.048919
std           0.014064          0.052813        0.079720
0.038803
min           0.052630          0.019380        0.000000
0.000000
```

```
25%          0.086370         0.064920        0.029560
0.020310
50%          0.095870         0.092630        0.061540
0.033500
75%          0.105300         0.130400        0.130700
0.074000
max          0.163400         0.345400        0.426800
0.201200

       mean symmetry  mean fractal dimension  ...  worst texture  \
count     569.000000              569.000000  ...     569.000000
mean        0.181162                0.062798  ...      25.677223
std         0.027414                0.007060  ...       6.146258
min         0.106000                0.049960  ...      12.020000
25%         0.161900                0.057700  ...      21.080000
50%         0.179200                0.061540  ...      25.410000
75%         0.195700                0.066120  ...      29.720000
max         0.304000                0.097440  ...      49.540000

       worst perimeter   worst area  worst smoothness  worst
compactness  \
count       569.000000   569.000000        569.000000
569.000000
mean        107.261213   880.583128          0.132369
0.254265
std          33.602542   569.356993          0.022832
0.157336
min          50.410000   185.200000          0.071170
0.027290
25%          84.110000   515.300000          0.116600
0.147200
50%          97.660000   686.500000          0.131300
0.211900
75%         125.400000  1084.000000          0.146000
0.339100
max         251.200000  4254.000000          0.222600
1.058000

       worst concavity  worst concave points  worst symmetry  \
count       569.000000            569.000000      569.000000
mean          0.272188              0.114606        0.290076
std           0.208624              0.065732        0.061867
min           0.000000              0.000000        0.156500
25%           0.114500              0.064930        0.250400
50%           0.226700              0.099930        0.282200
75%           0.382900              0.161400        0.317900
max           1.252000              0.291000        0.663800

       worst fractal dimension      target
count               569.000000  569.000000
```

```
mean                      0.083946    0.627417
std                       0.018061    0.483918
min                       0.055040    0.000000
25%                       0.071460    0.000000
50%                       0.080040    1.000000
75%                       0.092080    1.000000
max                       0.207500    1.000000

[8 rows x 31 columns]
```

```
df_cancer.shape
```

```
(569, 31)
```

Performing the Data Cleaning

```
df_cancer.duplicated().sum()
```

```
0
```

```
df_cancer.isnull().sum()
```

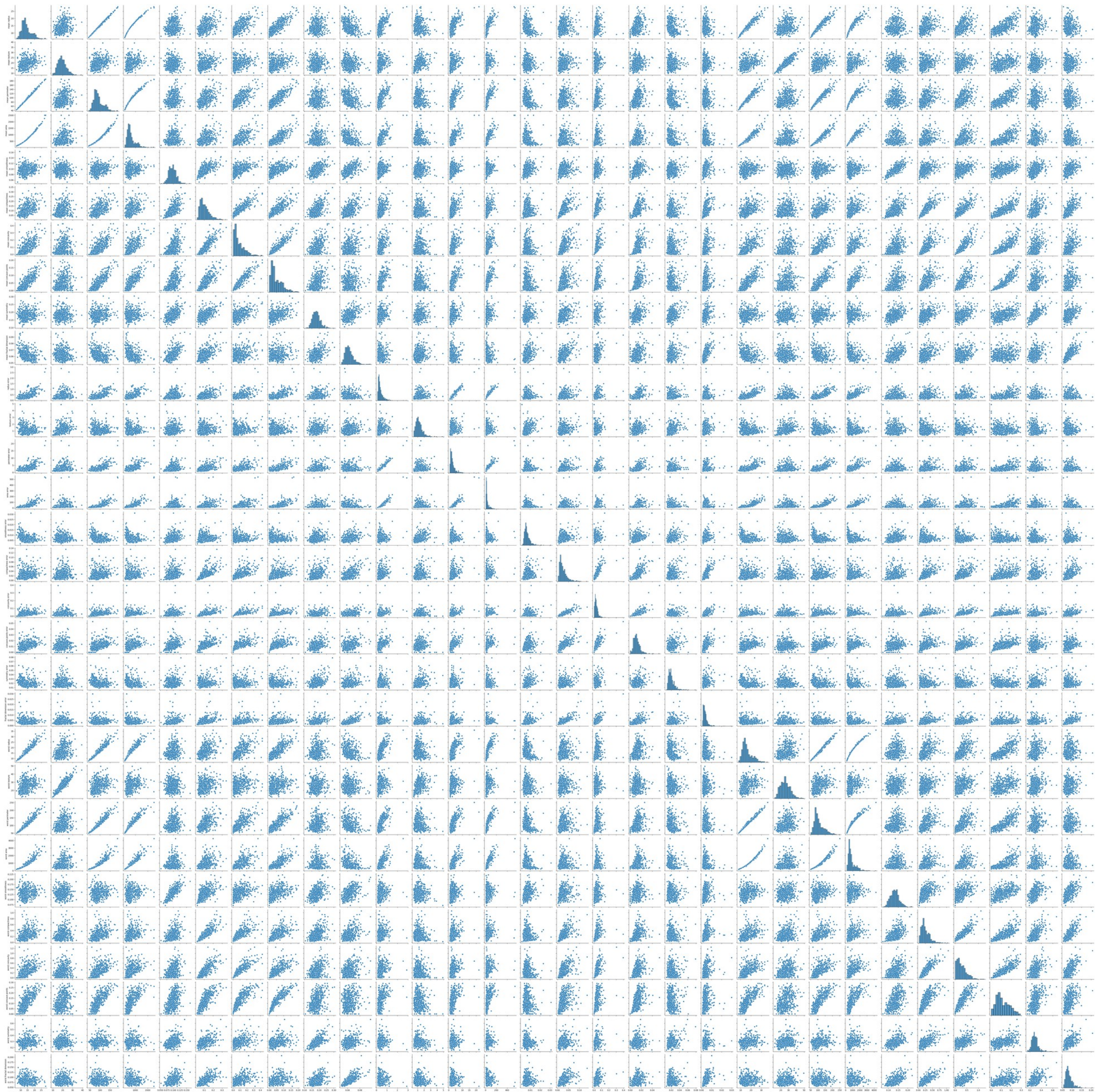
```
mean radius                0
mean texture               0
mean perimeter             0
mean area                  0
mean smoothness            0
mean compactness           0
mean concavity             0
mean concave points        0
mean symmetry              0
mean fractal dimension     0
radius error               0
texture error              0
perimeter error            0
area error                 0
smoothness error           0
compactness error          0
concavity error            0
concave points error       0
symmetry error             0
fractal dimension error    0
worst radius               0
worst texture              0
worst perimeter            0
worst area                 0
worst smoothness           0
worst compactness          0
worst concavity            0
worst concave points       0
```

```
worst symmetry             0
worst fractal dimension    0
target                     0
dtype: int64
```

Visualising the Dataset

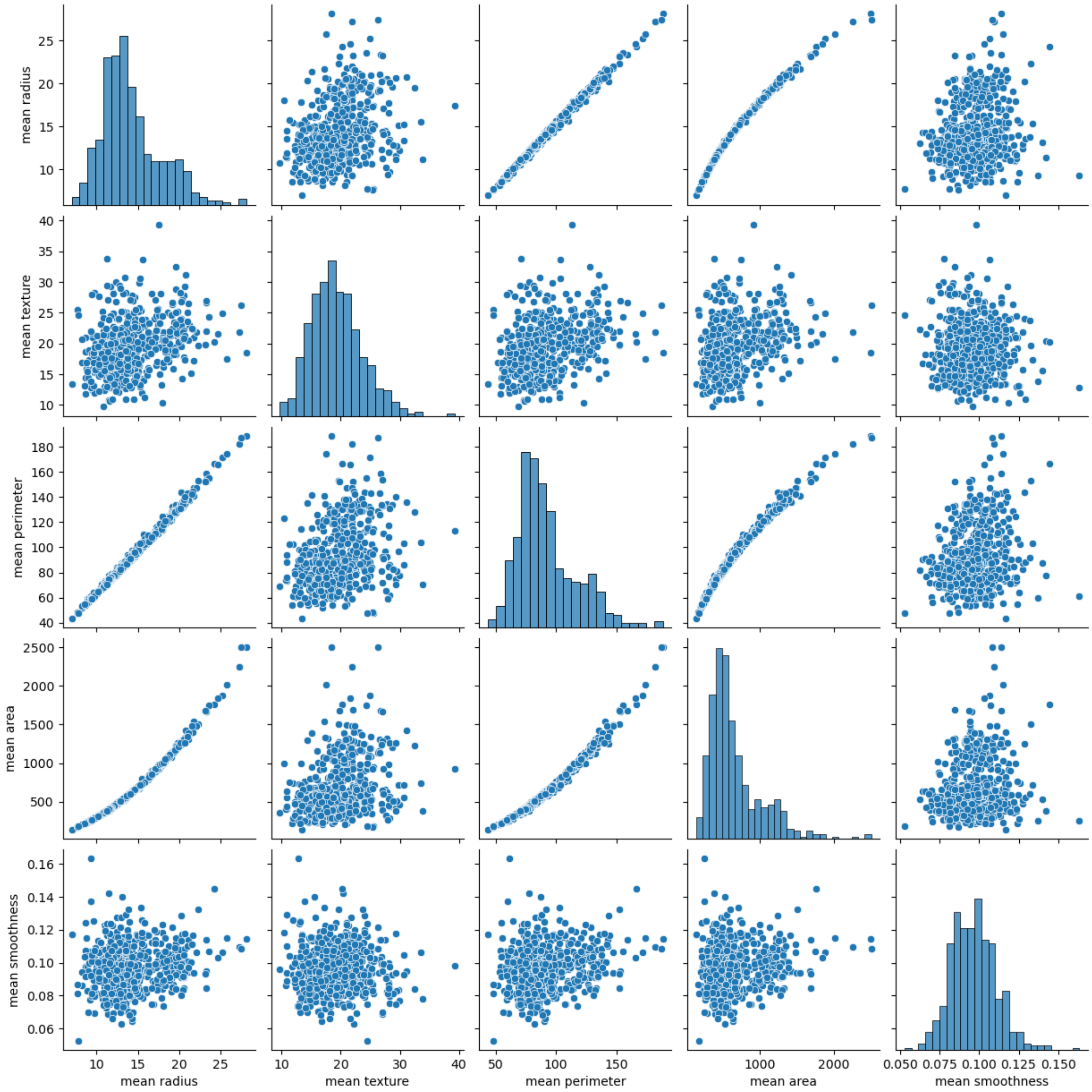
```
sns.pairplot(df_cancer , vars =['mean radius','mean texture', 'mean
perimeter', 'mean area',
 'mean smoothness', 'mean compactness' ,'mean concavity',
 'mean concave points', 'mean symmetry', 'mean fractal dimension',
 'radius error', 'texture error', 'perimeter error', 'area error',
 'smoothness error', 'compactness error', 'concavity error',
 'concave points error', 'symmetry error', 'fractal dimension error',
 'worst radius', 'worst texture', 'worst perimeter', 'worst area',
 'worst smoothness', 'worst compactness', 'worst concavity',
 'worst concave points', 'worst symmetry', 'worst fractal dimension'])
```

```
<seaborn.axisgrid.PairGrid at 0x15545ce69d0>
```



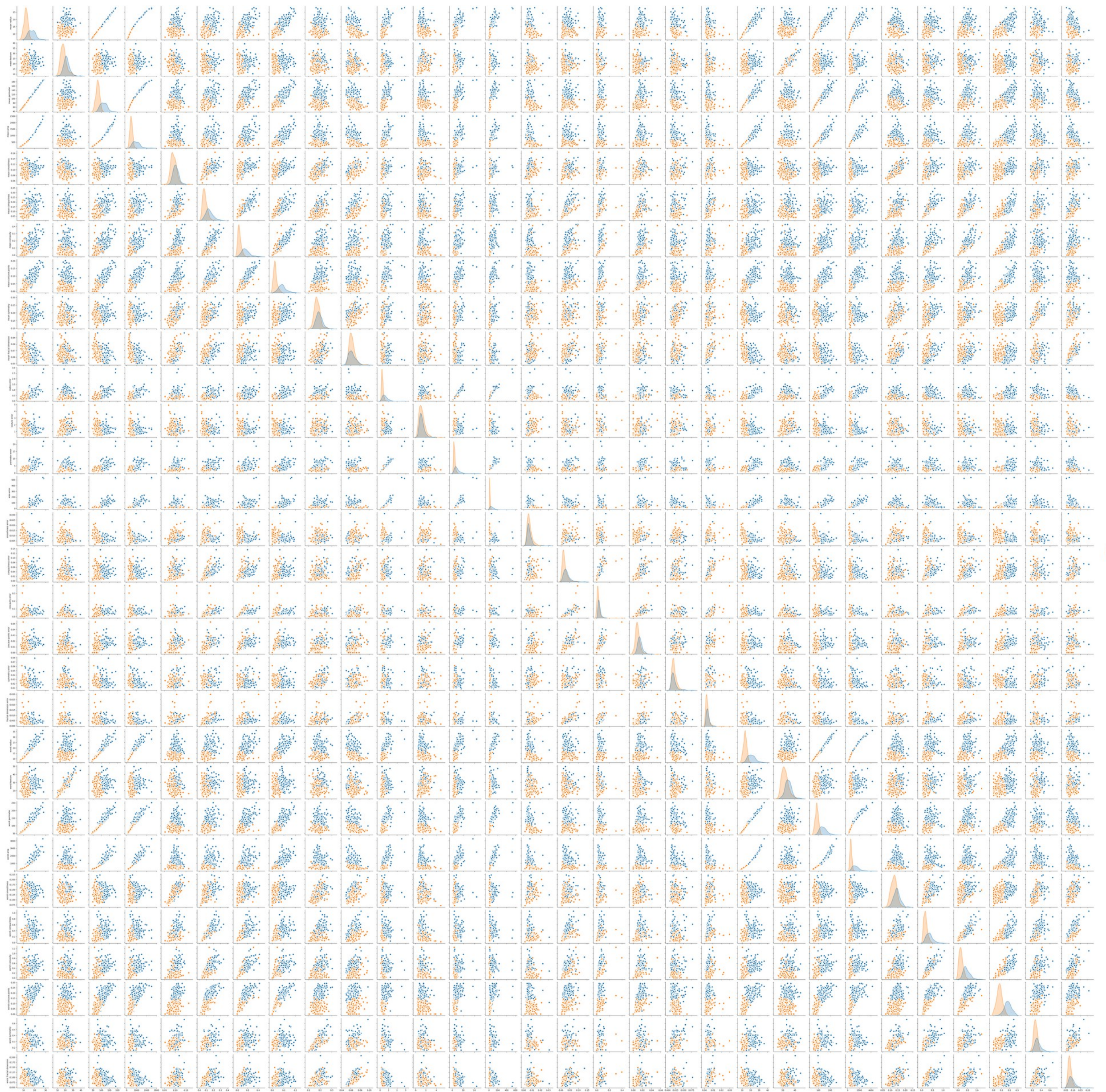
```
#taking only 5 variable out of 30 just to showcase how powerfull
seaborn library actually is
sns.pairplot(df_cancer, vars = ['mean radius', 'mean texture', 'mean
perimeter', 'mean area',
 'mean smoothness'])
```

```
<seaborn.axisgrid.PairGrid at 0x1557612e490>
```

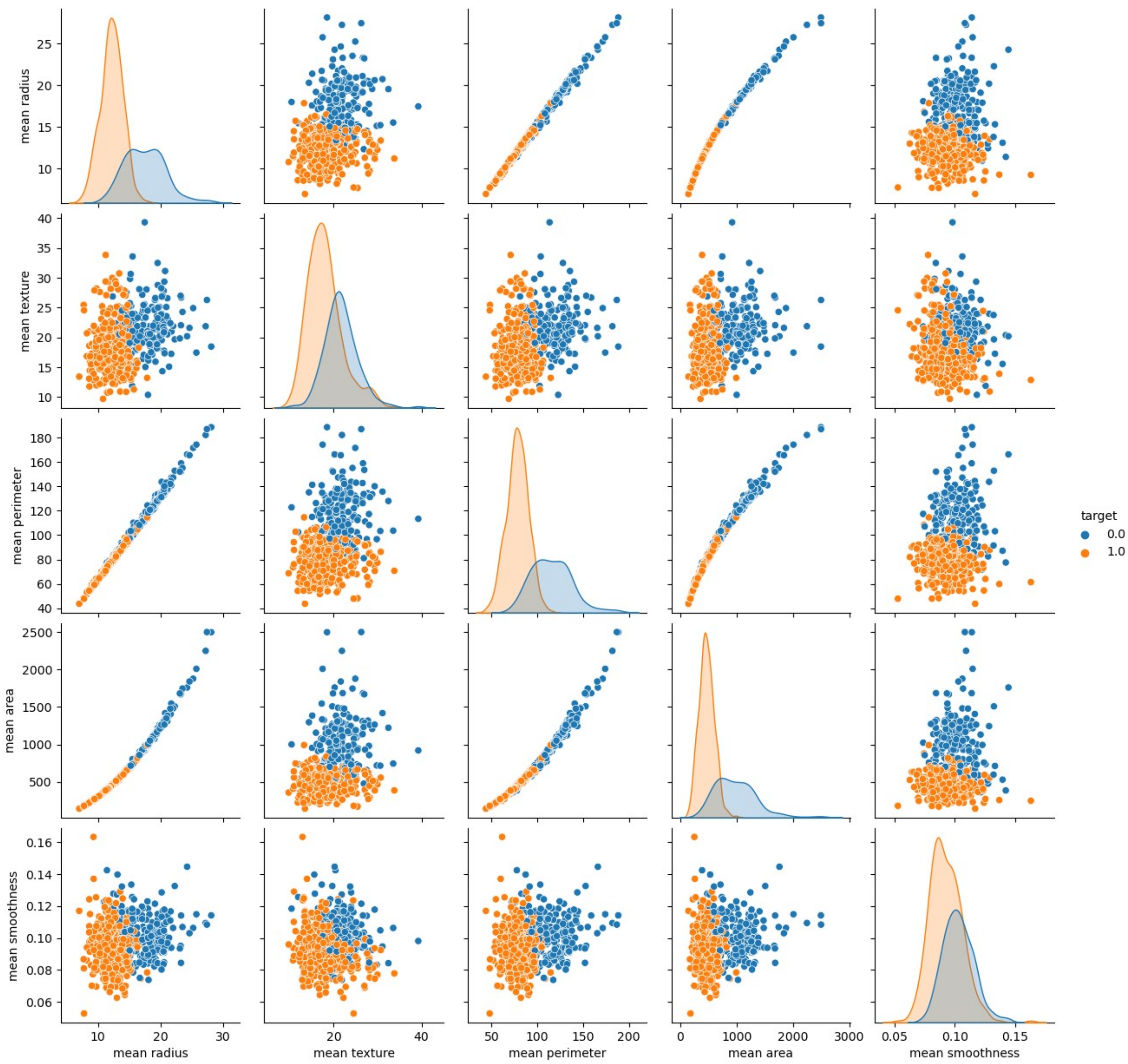



```
plt.figure(figsize=(10,8), dpi= 80)
sns.pairplot(df_cancer, kind="scatter", hue="target",
plot_kws=dict(s=80, edgecolor="white", linewidth=2.5))
plt.show()
```

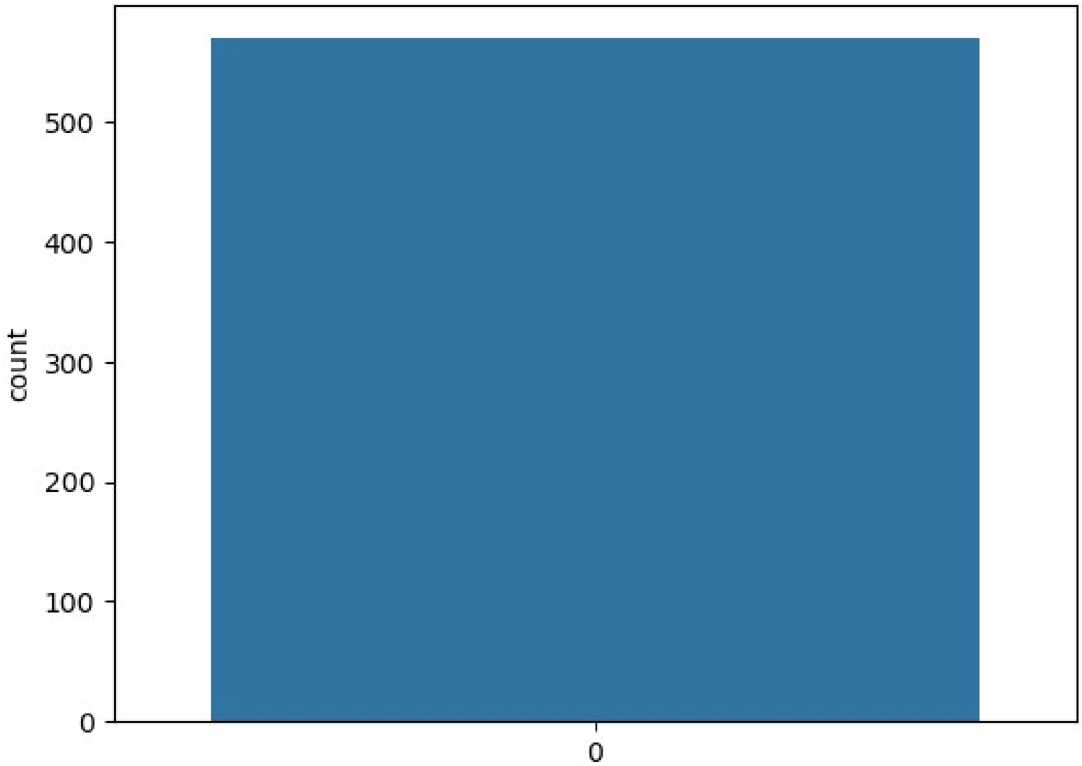
```
<Figure size 800x640 with 0 Axes>
```



```
#in above plotting we are not able to differentiate much,so we use
'hue' on target column, which will seperate the two(Malignent,
Benign).
sns.pairplot(df_cancer, hue = 'target', vars = ['mean radius', 'mean
texture', 'mean perimeter', 'mean area',
 'mean smoothness'])

#blue points are malignent case which are severe cases or life
threatning cases.
#orange points are not very severe or life threatning

<seaborn.axisgrid.PairGrid at 0x1557cfe4d90>
```

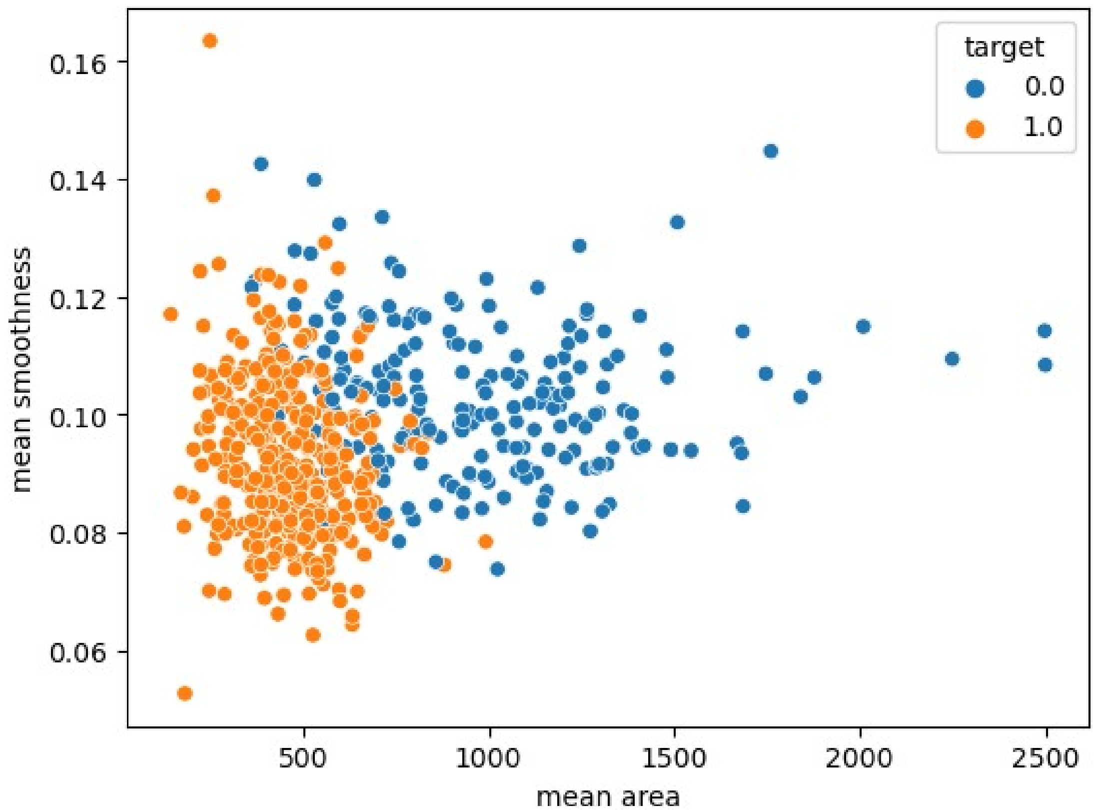



```
#will simply tell you how many Malignent and Benign cases we have.
#Malignent= 200~ and Benign = 350~ approx.
sns.countplot(df_cancer['target'])

<Axes: ylabel='count'>
```



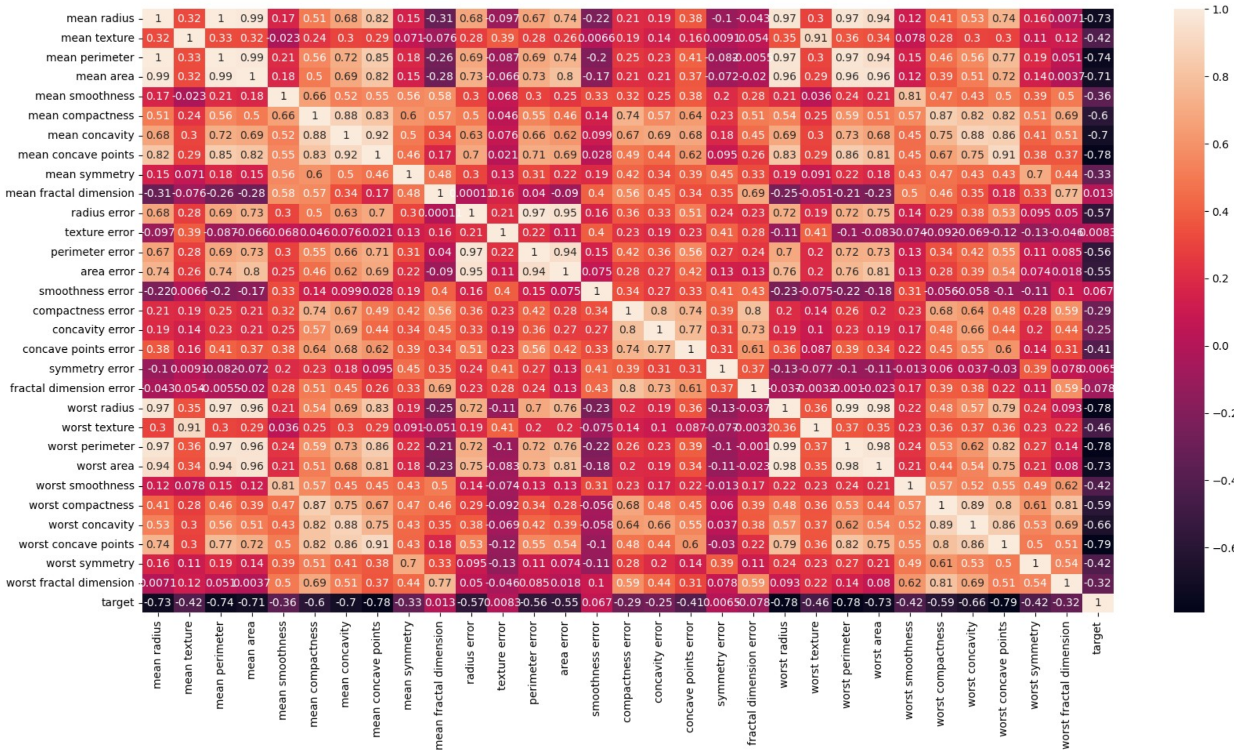

```python
#plotting a scatter plot diagram for mean area anf mean smoothness,
you can plot any feature combination scatterplot.
sns.scatterplot(x='mean area',y='mean smoothness',hue='target',data
=df_cancer)
```

```
<Axes: xlabel='mean area', ylabel='mean smoothness'>
```




```
#here we made a heatmap figure of correlation of all the columns
plt.figure(figsize =(20,10))
sns.heatmap(df_cancer.corr(), annot =True)
```

```
<Axes: >
```



Splitting the Dataset

```
x = df_cancer.drop(['target'],axis =1)
#train test split

x

     mean radius  mean texture  mean perimeter  mean area  mean
smoothness  \
0          17.99         10.38          122.80     1001.0
0.11840
1          20.57         17.77          132.90     1326.0
0.08474
2          19.69         21.25          130.00     1203.0
0.10960
3          11.42         20.38           77.58      386.1
0.14250
4          20.29         14.34          135.10     1297.0
0.10030
..           ...           ...             ...        ...
...
564        21.56         22.39          142.00     1479.0
0.11100
565        20.13         28.25          131.20     1261.0
0.09780
566        16.60         28.08          108.30      858.1
0.08455
```

```
567        20.60         29.33          140.10     1265.0
0.11780
568         7.76         24.54           47.92      181.0
0.05263

     mean compactness  mean concavity  mean concave points  mean
symmetry  \
0             0.27760         0.30010              0.14710
0.2419
1             0.07864         0.08690              0.07017
0.1812
2             0.15990         0.19740              0.12790
0.2069
3             0.28390         0.24140              0.10520
0.2597
4             0.13280         0.19800              0.10430
0.1809
..                ...             ...                  ...
...
564           0.11590         0.24390              0.13890
0.1726
565           0.10340         0.14400              0.09791
0.1752
566           0.10230         0.09251              0.05302
0.1590
567           0.27700         0.35140              0.15200
0.2397
568           0.04362         0.00000              0.00000
0.1587

     mean fractal dimension  ...  worst radius  worst texture  \
0                   0.07871  ...        25.380          17.33
1                   0.05667  ...        24.990          23.41
2                   0.05999  ...        23.570          25.53
3                   0.09744  ...        14.910          26.50
4                   0.05883  ...        22.540          16.67
..                      ...  ...           ...            ...
564                 0.05623  ...        25.450          26.40
565                 0.05533  ...        23.690          38.25
566                 0.05648  ...        18.980          34.12
567                 0.07016  ...        25.740          39.42
568                 0.05884  ...         9.456          30.37

     worst perimeter  worst area  worst smoothness  worst compactness
\
0             184.60      2019.0           0.16220            0.66560

1             158.80      1956.0           0.12380            0.18660

2             152.50      1709.0           0.14440            0.42450
```

```
3              98.87       567.7                0.20980                0.86630
4             152.20      1575.0                0.13740                0.20500
..               ...         ...                    ...                    ...
564           166.10      2027.0                0.14100                0.21130
565           155.00      1731.0                0.11660                0.19220
566           126.70      1124.0                0.11390                0.30940
567           184.60      1821.0                0.16500                0.86810
568            59.16       268.6                0.08996                0.06444

     worst concavity  worst concave points  worst symmetry  \
0             0.7119                0.2654          0.4601
1             0.2416                0.1860          0.2750
2             0.4504                0.2430          0.3613
3             0.6869                0.2575          0.6638
4             0.4000                0.1625          0.2364
..               ...                   ...             ...
564           0.4107                0.2216          0.2060
565           0.3215                0.1628          0.2572
566           0.3403                0.1418          0.2218
567           0.9387                0.2650          0.4087
568           0.0000                0.0000          0.2871

     worst fractal dimension
0                    0.11890
1                    0.08902
2                    0.08758
3                    0.17300
4                    0.07678
..                       ...
564                  0.07115
565                  0.06637
566                  0.07820
567                  0.12400
568                  0.07039

[569 rows x 30 columns]
```

```
y= df_cancer['target']

y
```

```
0      0.0
1      0.0
2      0.0
3      0.0
4      0.0
      ...
564    0.0
565    0.0
566    0.0
567    0.0
568    1.0
Name: target, Length: 569, dtype: float64
```

```
#'test_size' is what is the size of the test data whicg is 15% of the whole data. we had 569 rows which will get split by train = 483 and test = 86.
#look below for better understanding, we've printed all the 4 values for X_train, X_text, y_train, y_test.
#also make sure the X is capital and y is small..
from sklearn.model_selection import train_test_split

 x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.2, random_state=5)

 x_train
```

```
     mean radius  mean texture  mean perimeter  mean area  mean smoothness  \
306       13.200         15.82           84.07      537.3
0.08511
410       11.360         17.57           72.49      399.8
0.08858
197       18.080         21.84          117.40     1024.0
0.07371
376       10.570         20.22           70.15      338.3
0.09073
244       19.400         23.50          129.10     1155.0
0.10270
..           ...           ...             ...        ...
...
8         13.000         21.82           87.50      519.8
0.12730
73        13.800         15.79           90.43      584.1
0.10070
400       17.910         21.02          124.40      994.0
0.12300
118       15.780         22.91          105.70      782.6
0.11550
206        9.876         17.27           62.92      295.4
0.10890
```

```
     mean compactness  mean concavity  mean concave points  mean
symmetry  \
306           0.05251        0.001461             0.003261
0.1632
410           0.05313        0.027830             0.021000
0.1601
197           0.08642        0.110300             0.057780
0.1770
376           0.16600        0.228000             0.059410
0.2188
244           0.15580        0.204900             0.088860
0.1978
..                ...             ...                  ...
...
8             0.19320        0.185900             0.093530
0.2350
73            0.12800        0.077890             0.050690
0.1662
400           0.25760        0.318900             0.119800
0.2113
118           0.17520        0.213300             0.094790
0.2096
206           0.07232        0.017560             0.019520
0.1934

     mean fractal dimension  ...  worst radius  worst texture  \
306                 0.05894  ...         14.41          20.45
410                 0.05913  ...         13.05          36.32
197                 0.05340  ...         19.76          24.70
376                 0.08450  ...         10.85          22.82
244                 0.06000  ...         21.65          30.53
..                      ...  ...           ...            ...
8                   0.07389  ...         15.49          30.73
73                  0.06566  ...         16.57          20.86
400                 0.07115  ...         20.80          27.78
118                 0.07331  ...         20.19          30.50
206                 0.06285  ...         10.42          23.22

     worst perimeter  worst area  worst smoothness  worst compactness
\
306            92.00       636.9           0.11280             0.1346

410            85.07       521.3           0.14530             0.1622

197           129.10      1228.0           0.08822             0.1963

376            76.51       351.9           0.11430             0.3619

244           144.90      1417.0           0.14630             0.2968
```

```
..              ...        ...               ...              ...
8            106.20      739.3           0.17030           0.5401
73           110.30      812.4           0.14110           0.3542
400          149.60     1304.0           0.18730           0.5917
118          130.30     1272.0           0.18550           0.4925
206           67.08      331.6           0.14150           0.1247

     worst concavity  worst concave points  worst symmetry  \
306          0.01120               0.02500          0.2651
410          0.18110               0.08698          0.2973
197          0.25350               0.09181          0.2369
376          0.60300               0.14650          0.2597
244          0.34580               0.15640          0.2920
..               ...                   ...             ...
8            0.53900               0.20600          0.4378
73           0.27790               0.13830          0.2589
400          0.90340               0.19640          0.3245
118          0.73560               0.20340          0.3274
206          0.06213               0.05588          0.2989

     worst fractal dimension
306                  0.08385
410                  0.07745
197                  0.06558
376                  0.12000
244                  0.07614
..                       ...
8                    0.10720
73                   0.10300
400                  0.11980
118                  0.12520
206                  0.07380

[455 rows x 30 columns]
```

y_train

```
306    1.0
410    1.0
197    0.0
376    1.0
244    0.0
      ...
8      0.0
```

```
73     0.0
400    0.0
118    0.0
206    1.0
Name: target, Length: 455, dtype: float64
```

x_test

```
     mean radius  mean texture  mean perimeter  mean area  mean
smoothness  \
28         15.30         25.27          102.40      732.4
0.10820
163        12.34         22.22           79.85      464.5
0.10120
123        14.50         10.89           94.28      640.7
0.11010
361        13.30         21.57           85.24      546.1
0.08582
549        10.82         24.21           68.89      361.6
0.08192
..           ...           ...             ...        ...
...
414        15.13         29.81           96.71      719.5
0.08320
515        11.34         18.61           72.76      391.2
0.10490
186        18.31         18.58          118.60     1041.0
0.08588
3          11.42         20.38           77.58      386.1
0.14250
261        17.35         23.06          111.00      933.1
0.08662

     mean compactness  mean concavity  mean concave points  mean
symmetry  \
28            0.16970         0.16830              0.08751
0.1926
163           0.10150         0.05370              0.02822
0.1551
123           0.10990         0.08842              0.05778
0.1856
361           0.06373         0.03344              0.02424
0.1815
549           0.06602         0.01548              0.00816
0.1976
..                ...             ...                  ...
...
414           0.04605         0.04686              0.02739
0.1852
515           0.08499         0.04302              0.02594
```

```
0.1927
186          0.08468         0.08169              0.05814
0.1621
3            0.28390         0.24140              0.10520
0.2597
261          0.06290         0.02891              0.02837
0.1564

     mean fractal dimension  ...  worst radius  worst texture  \
28                  0.06540  ...         20.27          36.71
163                 0.06761  ...         13.58          28.68
123                 0.06402  ...         15.70          15.98
361                 0.05696  ...         14.20          29.20
549                 0.06328  ...         13.03          31.45
..                      ...  ...           ...            ...
414                 0.05294  ...         17.26          36.91
515                 0.06211  ...         12.47          23.03
186                 0.05425  ...         21.31          26.36
3                   0.09744  ...         14.91          26.50
261                 0.05307  ...         19.85          31.47

     worst perimeter  worst area  worst smoothness  worst compactness
\
28            149.30      1269.0            0.1641            0.61100

163            87.36       553.0            0.1452            0.23380

123           102.80       745.5            0.1313            0.17880

361            92.94       621.2            0.1140            0.16670

549            83.90       505.6            0.1204            0.16330

..               ...         ...               ...                ...

414           110.10       931.4            0.1148            0.09866

515            79.15       478.6            0.1483            0.15740

186           139.20      1410.0            0.1234            0.24450

3              98.87       567.7            0.2098            0.86630

261           128.20      1218.0            0.1240            0.14860


     worst concavity  worst concave points  worst symmetry  \
28           0.63350               0.20240          0.4027
163          0.16880               0.08194          0.2268
123          0.25600               0.12210          0.2889
361          0.12120               0.05614          0.2637
```

```
549          0.06194               0.03264          0.3059
..               ...                   ...             ...
414          0.15470               0.06575          0.3233
515          0.16240               0.08542          0.3060
186          0.35380               0.15710          0.3206
3            0.68690               0.25750          0.6638
261          0.12110               0.08235          0.2452

     worst fractal dimension
28                   0.09876
163                  0.09082
123                  0.08006
361                  0.06658
549                  0.07626
..                       ...
414                  0.06165
515                  0.06783
186                  0.06938
3                    0.17300
261                  0.06515

[114 rows x 30 columns]
```

```
y_test
```

```
28     0.0
163    1.0
123    1.0
361    1.0
549    1.0
      ...
414    0.0
515    1.0
186    0.0
3      0.0
261    0.0
Name: target, Length: 114, dtype: float64
```

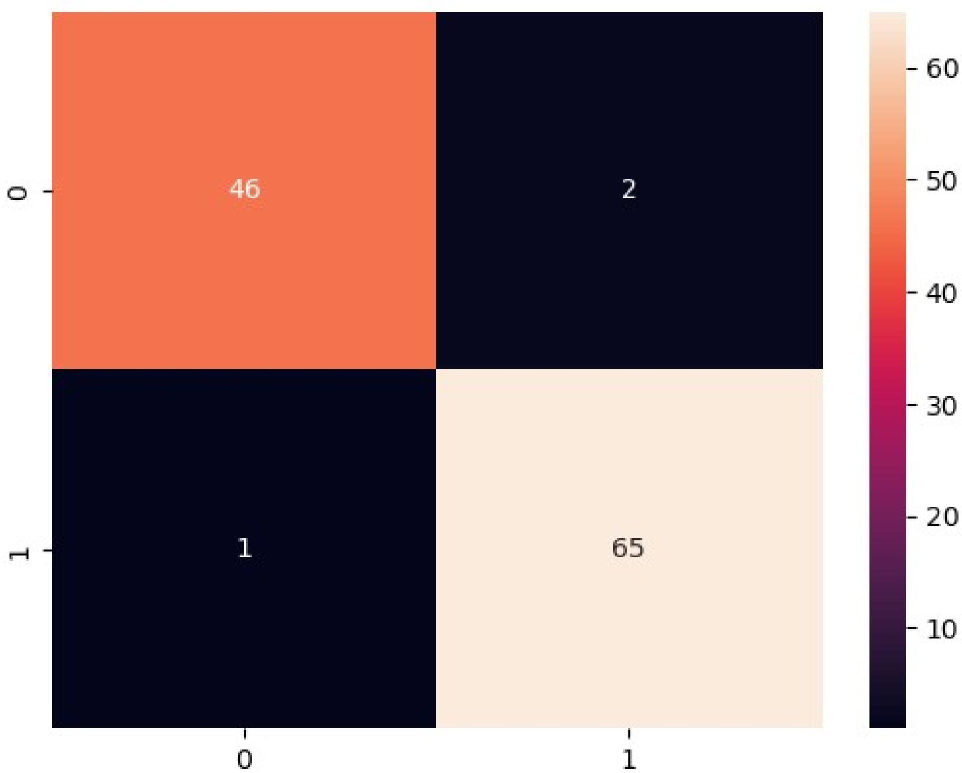
Training the Model using SVM (Support Vector Machine)

```
from sklearn.svm import SVC

from sklearn.metrics import classification_report, confusion_matrix

from sklearn.svm import SVC
svm_model = SVC(kernel = 'linear', random_state = 0)
svm_model.fit(x_train, y_train)
```

```
SVC(kernel='linear', random_state=0)
```

Evaluating the Model

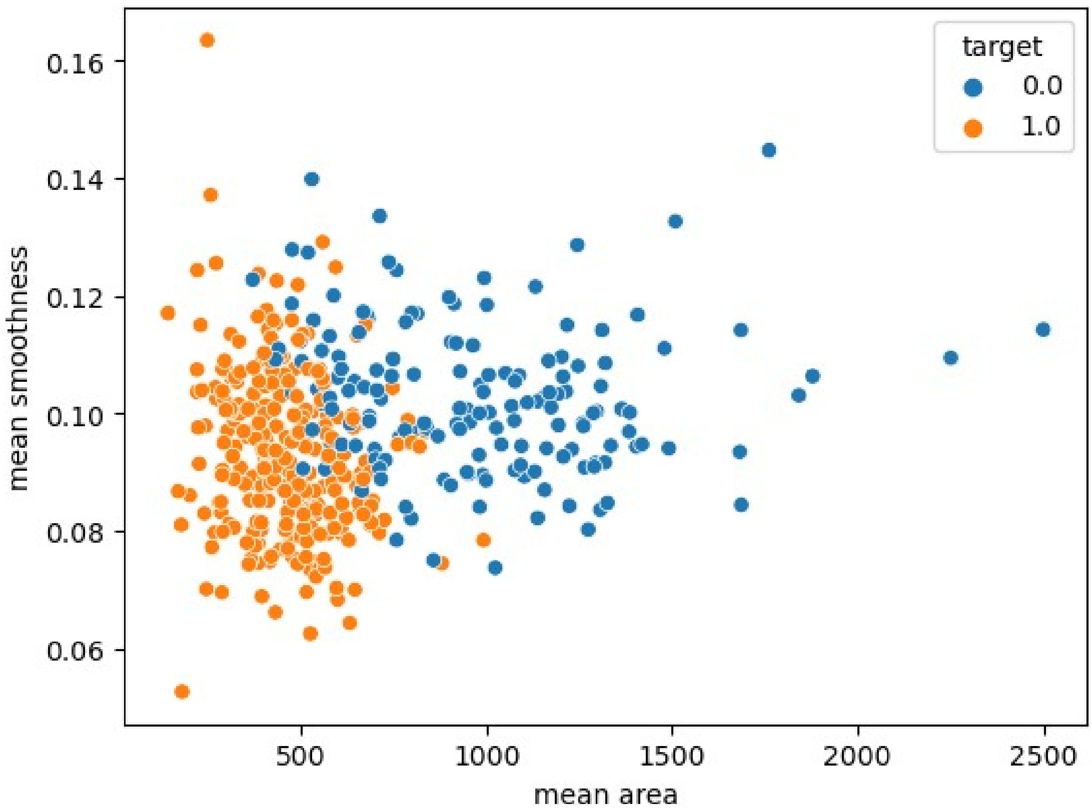
```
y_predict =svm_model.predict(x_test)

y_predict

array([0., 1., 1., 1., 1., 0., 1., 1., 1., 1., 1., 1., 0., 1., 1., 1.,
1.,
       1., 1., 1., 0., 1., 1., 1., 1., 1., 0., 0., 1., 0., 0., 0., 1.,
0.,
       1., 1., 0., 1., 1., 0., 1., 1., 1., 0., 1., 1., 0., 0., 1., 0.,
1.,
       1., 1., 1., 1., 0., 0., 0., 1., 0., 0., 0., 1., 1., 1., 1., 1.,
1.,
       1., 0., 1., 0., 1., 1., 1., 1., 0., 1., 0., 0., 0., 1., 0., 0.,
0.,
       1., 0., 1., 0., 0., 0., 0., 1., 1., 0., 0., 1., 1., 1., 1., 1.,
0.,
       1., 1., 0., 0., 0., 0., 1., 0., 1., 0., 0., 0.])

cm = confusion_matrix(y_test,y_predict)

sns.heatmap(cm ,annot=True)

<Axes: >
```



Model Improvisation

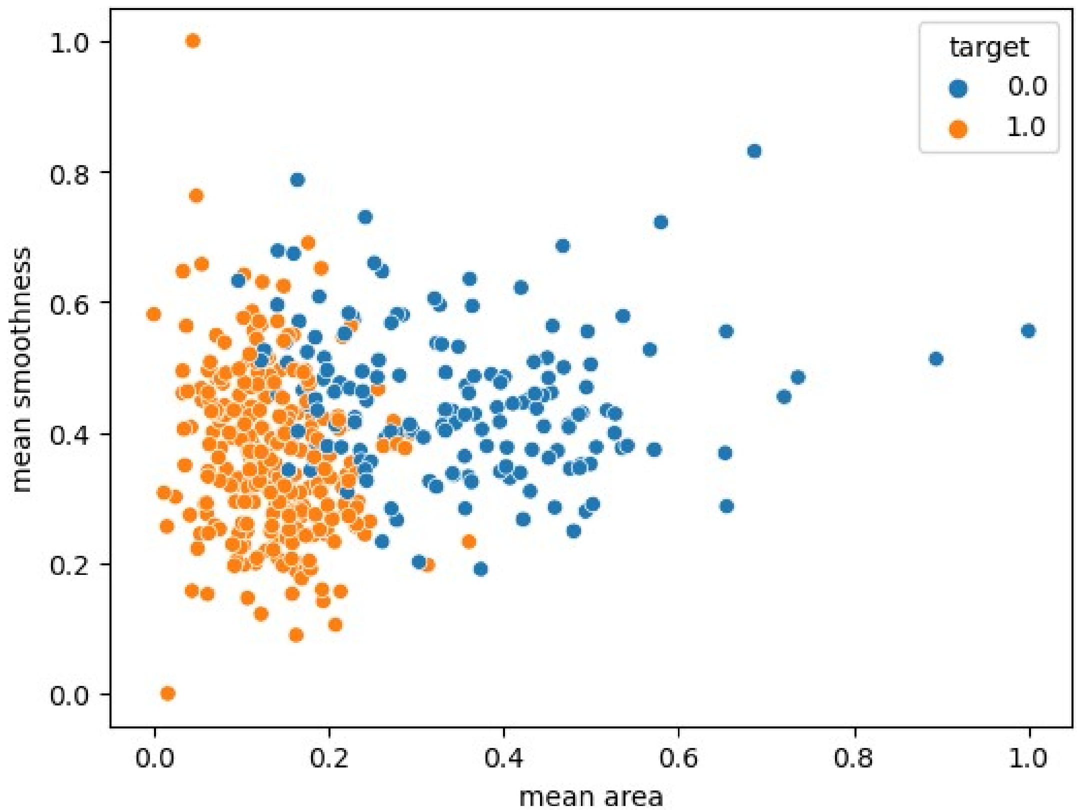
```
min_train =x_train.min()

range_train =(x_train - min_train).max()

x_train_scaled =(x_train-min_train)/range_train

sns.scatterplot(x = x_train['mean area'], y= x_train['mean smoothness'],hue =y_train)
```

```
<Axes: xlabel='mean area', ylabel='mean smoothness'>
```

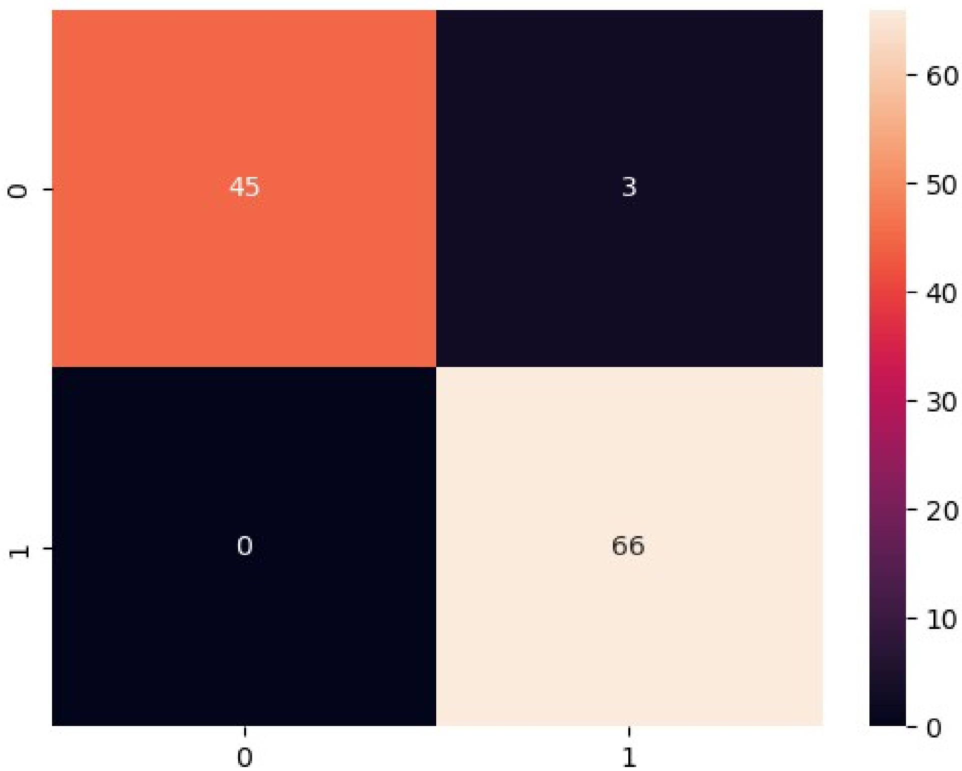



```
sns.scatterplot(x = x_train_scaled['mean area'], y=
x_train_scaled['mean smoothness'],hue =y_train)
```

```
<Axes: xlabel='mean area', ylabel='mean smoothness'>
```




```
min_test =x_test.min()
range_test =(x_test - min_test).max()
x_test_scaled =(x_test-min_test)/range_test

svm_model.fit(x_train_scaled,y_train)

SVC(kernel='linear', random_state=0)

y_predict =svm_model.predict(x_test_scaled)

cn = confusion_matrix(y_test,y_predict)

sns.heatmap(cn,  annot = True)

<Axes: >
```




```
print(classification_report(y_test,y_predict))

              precision    recall  f1-score   support

         0.0       1.00      0.94      0.97        48
         1.0       0.96      1.00      0.98        66

    accuracy                           0.97       114
   macro avg       0.98      0.97      0.97       114
weighted avg       0.97      0.97      0.97       114
```