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-011,
       [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-021.
       [2.060e+01, 2.933e+01, 1.401e+02, ..., 2.650e-01, 4.087e-01,
        1.240e-011,
       [7.760e+00, 2.454e+01, 4.792e+01, ..., 0.000e+00, 2.871e-01,
       7.039e-02]]),
0, 0, 1, 1, 1,
       0,
       0, 0, 1, 0, 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, 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, 0, 0, 1, 1, 0, 1, 1, 0, 0, 1, 0, 1,
1,
        1, 1, 0, 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, 0, 1, 0, 1, 1, 1, 1,
1,
       1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 0, 1, 0, 1, 1, 1, 1, 0,
0,
       0, 1, 1, 1, 1, 0, 1, 0, 1, 0, 1, 1, 1, 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, 1, 0,
0,
        1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 0, 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, 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**Data Set Characteristics:**\n\n :Number of Instances: 569\n\n
:Number of Attributes: 30 numeric, predictive attributes and the
             :Attribute Information:\n
class\n\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
```

```
- fractal dimension ("coastline approximation" - 1)
          The mean, standard error, and "worst" or largest (mean of
n n
the three\n
                  worst/largest values) of these features were
computed for each image,\n
                                resulting in 30 features.
instance, field 0 is Mean Radius, field\n
                                               10 is Radius SE,
                                                             - WDBC-
field 20 is Worst Radius.\n\n
                                   - class:\n
Malignant\n
                          - WDBC-Benign\n\n
                                              :Summary Statistics:\
n \ n
                      Min
      Max\n
               radius (mean):
                                    6.981
                                           28.11\n
                                                      texture
                             9.71
                                   39.28\n
                                              perimeter (mean):
(mean):
43.79 188.5\n
                 area (mean):
                                                      143.5 2501.0
                                         0.053 0.163\n
     smoothness (mean):
                                    0.019
                                           0.345\n
compactness (mean):
                                                      concavity
(mean):
                           0.0
                                 0.427\n
                                            concave points (mean):
      0.201\n
                                                      0.106 0.304\n
0.0
                 symmetry (mean):
fractal dimension (mean):
                                    0.05
                                           0.097\n
                                                      radius
(standard error):
                              0.112
                                    2.873\n
                                               texture (standard
                          4.885\n
error):
                   0.36
                                    perimeter (standard error):
                 area (standard error):
0.757 21.98\n
                                                      6.802 542.2\n
                                    0.002
smoothness (standard error):
                                           0.031\n
                                                      compactness
(standard error):
                         0.002 \quad 0.135\n
                                          concavity (standard
                                  concave points (standard error):
error):
                 0.0
                        0.396\n
0.0
      0.053\n
                 symmetry (standard error):
                                                      0.008 \quad 0.079 \ n
fractal dimension (standard error):
                                    0.001 \quad 0.03\n
                                                     radius (worst):
7.93
      36.04\n
                 texture (worst):
                                                      12.02 49.54\n
                                                      area (worst):
perimeter (worst):
                                    50.41 251.2\n
185.2 4254.0\n
                  smoothness (worst):
                                                       0.071 \quad 0.223
    compactness (worst):
                                         0.027 1.058\n
                                                           concavity
                                 1.252\n
                           0.0
(worst):
                                            concave points (worst):
                 symmetry (worst):
0.0
      0.291\n
                                                      0.156 \quad 0.664\n
fractal dimension (worst):
                                    0.055 \quad 0.208\n
:Class Distribution: 212 - Malignant,
Attribute Values: None\n\n
357 - Benian 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.ql/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
                  for breast tumor diagnosis. IS&T/SPIE 1993
extraction \n
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
Wolberg, W.N. Street, and O.L. Mangasarian. Machine learning
              to diagnose breast cancer from fine-needle aspirates.
techniques\n
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-021,
       [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-021,
      [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,
       0, 0, 0, 0, 0, 1, 1, 1,
      0,
      0, 0, 1, 0, 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, 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, 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, 0, 1, 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, 0, 1, 0, 1, 1, 1, 1,
1,
      1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 0, 1, 0, 1, 1, 1, 1, 0,
0,
      0, 1, 1, 1, 1, 0, 1, 0, 1, 0, 1, 1, 1, 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, 1, 0,
0,
      1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 0, 1, 1, 0, 1, 1, 0, 0, 1,
1,
      1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 0, 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, 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)
Characteristics:**\n\n :Number of Instances: 569\n\n
                                                         :Number of
Attributes: 30 numeric, predictive attributes and the class\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
- smoothness (local variation in radius lengths)\n
of concave portions of the contour)\n
compactness (perimeter^2 / area - 1.0)\n - concavity (severity
                                           - concave points (number
                                           symmetry\n
fractal dimension ("coastline approximation" - 1)\n\n
standard error, and "worst" or largest (mean of the three\n
worst/largest values) of these features were computed for each image,\
         resulting in 30 features. For instance, field 0 is Mean
                      10 is Radius SE, field 20 is Worst Radius.\n\n
Radius, field\n
- class:\n
                         - WDBC-Malignant\n - WDBC-
Benign\n\n
             :Summary Statistics:\n\n
               _____
                                               ==== =====\n
      Max\n
                                    6.981 28.11\n
radius (mean):
                                                      texture
                                   39.28\n
(mean):
                             9.71
                                              perimeter (mean):
43.79 188.5\n
                                                      143.5 2501.0
                 area (mean):
    smoothness (mean):
                                         0.053 0.163\n
compactness (mean):
                                    0.019 \quad 0.345 \ n
                                                      concavity
(mean):
                           0.0
                                 0.427\n
                                            concave points (mean):
0.0
       0.201\n
                 symmetry (mean):
                                                      0.106 \quad 0.304 \ n
fractal dimension (mean):
                                    0.05
                                           0.097\n
                                                      radius
(standard error):
                              0.112
                                    2.873\n
                                               texture (standard
                   0.36
                          4.885\n
                                     perimeter (standard error):
error):
                 area (standard error):
0.757 21.98\n
                                                      6.802 542.2\n
smoothness (standard error):
                                    0.002
                                           0.031\n
                                                      compactness
(standard error):
                         0.002 \quad 0.135\n
                                          concavity (standard
                                  concave points (standard error):
error):
                 0.0
                        0.396\n
0.0
      0.053\n
                 symmetry (standard error):
                                                      0.008 \quad 0.079 \ n
fractal dimension (standard error):
                                                     radius (worst):
                                    0.001 \quad 0.03\n
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\
                                         0.027 \quad 1.058\n
                                                          concavity
    compactness (worst):
(worst):
                                 1.252\n concave points (worst):
                           0.0
0.0
      0.291\n
                 symmetry (worst):
                                                      0.156 \quad 0.664 \ n
```

```
fractal dimension (worst):
                                     0.055 0.208\n
                     =======\n\n
                                                         :Missing
Attribute Values: None\n\n
                             :Class Distribution: 212 - Malignant,
357 - Benian 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
                 for breast tumor diagnosis. IS&T/SPIE 1993
extraction \n
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
                    prognosis via linear programming. Operations
diagnosis and \n
Research, 43(4), pages 570-577, \n
                                      July-August 1995.\n
Wolberg, W.N. Street, and O.L. Mangasarian. Machine learning
                to diagnose breast cancer from fine-needle aspirates.
techniques\n
Cancer Letters 77 (1994) n 163-171.', array(['mean radius', 'mean
'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\
     :Attribute Information:\n - radius (mean of distances from
center to points on the perimeter)\n - texture (
deviation of gray-scale values)\n - perimeter\n

    texture (standard

- smoothness (local variation in radius lengths)\n
compactness (perimeter^2 / area - 1.0)\n - concavity (severity
of concave portions of the contour)\n
of concave portions of the contour)\n

    concave points (number

                                           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,\
        resulting in 30 features. For instance, field 0 is Mean
                      10 is Radius SE, field 20 is Worst Radius.\n\n
Radius, field\n
- class:\n
                         - WDBC-Malignant\n
Benign\n\n :Summary Statistics:\n\n
Max\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
    smoothness (mean):
                                         0.053 \quad 0.163\n
compactness (mean):
                                    0.019 \quad 0.345 \ n
                                                     concavity
                          0.0
                                 0.427\n concave points (mean):
(mean):
      0.201\n
                 symmetry (mean):
                                                      0.106 \quad 0.304 \ n
0.0
                                           0.097\n
fractal dimension (mean):
                                    0.05
                                                      radius
(standard error):
                             0.112 2.873\n
                                             texture (standard
                   0.36
                         4.885\n
                                    perimeter (standard error):
error):
0.757 21.98\n
                area (standard error):
                                                     6.802 542.2\n
smoothness (standard error):
                                    0.002 \quad 0.031\n
                                                      compactness
(standard error):
                        0.002 \quad 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 \quad 0.079 \ n
fractal dimension (standard error):
                                    0.001 \quad 0.03\n
                                                     radius (worst):
      36.04\n
                texture (worst):
                                                     12.02 49.54\n
7.93
                                    50.41 251.2\n
perimeter (worst):
                                                     area (worst):
185.2 4254.0\n
                  smoothness (worst):
                                                      0.071 0.223\
    compactness (worst):
                                         0.027 \quad 1.058\n
                                                          concavity
                                 1.252\n concave points (worst):
(worst):
                           0.0
0.0
                 symmetry (worst):
                                                     0.156 \quad 0.664\n
      0.291\n
                                    0.055 0.208\n
fractal dimension (worst):
==========\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 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.'

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																																	
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	_		_	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_			_		_	_	_	_	_	_
$\frac{1}{1}$	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	Τ	Τ	Τ	0	Τ	Ι	0	0	Τ	0	0	0	0	Τ	0	0	0	Τ	0	Τ	0	Τ	Τ	0	1	0	0	0	0	Τ	Ι	0	0
1 1	-	-	-	-	-	^	^	-	-	^	-	1	^	^	-	^	-	-	-	-	^	1	-	1	1	-	^	1	^	^	^	^	^
$\frac{1}{0}$	Т	T	Τ	Т	Τ	U	O	Τ	Τ	0	Τ	Τ	0	O	Τ	0	Т	Τ	Τ	Τ	0	Τ	Т	Τ	Τ	Τ	U	Τ	0	0	U	0	0
0 0	Λ	^	Λ	^	Λ	1	1	1	1	1	1	Λ	1	^	1	1	^	1	1	Λ	1	Λ	^	1	1	1	1	1	1	1	1	1	7
$egin{array}{ccc} 0 & 0 \ 1 & 1 \end{array}$	U	U	U	U	U	T	Т	Т	Т	Т	Т	U	Т	U	Т	Т	U	Т	Т	U	Т	U	U	Т	Τ	Т	Т	Т	Т	1	T	1	
1 0	1	1	a	1	a	1	1	1	1	1	1	1	1	1	1	1	1	1	1	O	1	1	1	۵	1	a	1	1	1	1	۵	0	a
1 1			U		U	1		_	_			_	_		_			_	_	U	_	_		U	_	U	1	_		_	U	U	U
1 1	0	1	0	1	O.	1	1	1	0	1	1	1	1	1	1	1	0	O	0	1	1	1	1	1	1	1	1	1	1	1	O	0	1
0 0	U	_	U	_	U	_	_	_	U	_	_	_	_	_	_	_	U	U	U	_	_	_	_	_	_	_	_	_	_	_	U	U	_
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 1
1 1 1 1 1 1 0 1 0 1 1 1 1 1 1 1 1 0 0 1 0 1 0 1 1 1 1 1 1 0 1 1 0 1 1
0 0
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 \
                                   122.80
         17.99
                     10.38
                                            1001.0
0.11840
         20.57
                     17.77
                                   132.90
                                            1326.0
1
0.08474
                     21.25
         19.69
                                   130.00
                                            1203.0
0.10960
                     20.38
                                             386.1
         11.42
                                    77.58
0.14250
         20.29
                      14.34
                                   135.10
                                            1297.0
0.10030
. . .
         21.56
564
                      22.39
                                   142.00
                                            1479.0
0.11100
565
                     28.25
         20.13
                                   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	20.00	29.33	140.10	1205.0	
568	7.76	24.54	47.92	181.0	
0.05263					
	•	mean conca	vity mean con	cave points	mean
symmetry	0.27760	0.2	0010	0 14710	
0 0.2419	0.27760	0.3	0010	0.14710	
1	0.07864	0.0	8690	0.07017	
0.1812	0.07004	0.0	0030	0.07017	
2	0.15990	0.1	9740	0.12790	
0.2069					
3	0.28390	0.2	4140	0.10520	
0.2597					
4	0.13280	0.1	9800	0.10430	
0.1809					
• •					
 564	0.11590	0.2	4390	0.13890	
0.1726	0.11330	012	1550	0.15050	
565	0.10340	0.1	4400	0.09791	
0.1752					
566	0.10230	0.0	9251	0.05302	
0.1590	0 07700		E1.40	0 15000	
567	0.27700	0.3	5140	0.15200	
0.2397 568	0.04362	0.0	0000	0.00000	
0.1587	0.04302	0.0	0000	0.00000	
0.1507					
mean	fractal dime	ension	worst texture	worst perim	eter
worst area					
0	0.	07871	17.33	18	4.60
2019.0 1	0	05667	22 /1	16	8.80
1956.0	υ.	05667	23.41	. 13	0.00
2	Θ.	05999	25.53	15	2.50
1709.0	0.		23.33		
3	0.	09744	26.50	9	8.87
567.7					
4	0.	05883	16.67	15	2.20
1575.0					
• •					
 564	O.	05623	26.40	16	6.10
2027.0	0.	05025	20.40	10	5.10

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		0.03004		30.37	39.10
20010					
	rst smoothnes		compactnes		
0	0.1622		0.6656		9.7119
1 2 3	0.1238 0.1444		0.1866 0.4245		9.2416 9.4504
3	0.2098		0.8663		0.6869
4	0.1374		0.2050		9.4000
564	0.1410		0.2113		9.4107
565 566	0.1166 0.1139		0.1922 0.3094		0.3215 0.3403
567	0.1650		0.8681		9.9387
568	0.0899		0.0644		0.0000
	rst concave p	oints wo	rst symmet	ry worst fr	actal dimension
target 0	G	.2654	0.46	0 1	0.11890
0.0	C	.2054	0.40	01	0.11090
1	e	.1860	0.27	50	0.08902
0.0					
2	e	.2430	0.36	13	0.08758
0.0 3	C	. 2575	0.66	30	0.17300
0.0	C	.2373	0.00	30	0.17300
4	e	.1625	0.23	64	0.07678
0.0					
 564	C	.2216	0.20	60	0.07115
564 0.0	C	.2210	0.20	00	0.0/113
565	e	.1628	0.25	72	0.06637
0.0					
566	e	.1418	0.22	18	0.07820
0.0		2650	0 40	07	0 12400
567 0.0	E	. 2650	0.40	0/	0.12400
568	e	.0000	0.28	71	0.07039
1.0			0.20		
[569 ro	ws x 31 colum	ins]			

Performing Data Preprocessing

df_cancer.head() mean radius smoothness \ 0 mean texture mean perimeter mean area mean smoothness \ 0 mean 17.99 10.38 122.80 1001.0 0.11840 1 20.57 17.77 132.90 1326.0 0.08474 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 0.2609 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 0 0.05999 25.53 152.50 1709.0 3 0.09744	Perform	iing Dai	laPre	proces	sing						
smoothness () 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 0.2597 4 0.13280 0.1980 0.10430 0.1809 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 <	df_cance	r.head()		-						
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 0.07871 17.33 184.60 2019.0 1 0.05667 23.41 158.80 1956.0 2 1 0.05999 25.53 152.50 1709.0 3 0.09744 26.50 98.87 567.7 4 0.05883 16.67 152.20 worst smoothness worst compactness worst concavity worst concave points \ 0 0 0.1622 0.6656 0.7119 0.2654 1 0.1238 0.1866 0.2416			mean	textur	e meau	n peri	meter	mean a	rea	mean	
1	0			10.3	8	12	22.80	100	01.0		
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 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 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 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	1	20.57		17.7	7	13	32.90	132	26.0		
11.42	2	19.69		21.2	5	13	30.00	120	03.0		
4 20.29 14.34 135.10 1297.0 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 0.2669 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	3	11.42		20.3	8	<u>.</u>	77.58	38	86.1		
symmetry \	4	20.29		14.3	4	13	35.10	129	7.0		
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		-	ness	mean c	oncavi	ty mea	an con	cave po	ints	mean	
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 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	0		7760		0.30	91		0.1	4710		
2	1	0.0	7864		0.08	69		0.0	7017		
3	2	0.1	5990		0.19	74		0.1	2790		
4 0.13280 0.1980 0.10430 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	3	0.2	8390		0.24	14		0.1	.0520		
mean fractal dimension worst texture worst perimeter worst area \ 0	4	0.1	3280		0.198	30		0.1	.0430		
area \ 0		fractal	dimon	scion	\. <i>u</i>	orst ta	ovturo	word	nor	imotor	worst
2019.0 1	area \	TTACLAL			w	טואנ ני			·		WOTSC
1956.0 2	2019.0										
1709.0 3	1956.0										
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	1709.0										
1575.0 worst smoothness worst compactness worst concavity worst concave points \ 0	567.7										
points \ 0 0.1622 0.6656 0.7119 0.2654 1 0.1238 0.1866 0.2416 0.1860			0.0)5883			16.67			152.20	
0 0.1622 0.6656 0.7119 0.2654 0.1238 0.1866 0.2416 0.1860 0.2416			ness	worst	compac ⁻	tness	worst	concav	'ity	worst	concave
1 0.1238 0.1866 0.2416 0.1860	0	-	1622		0	. 6656		0.7	119		
	1	0.	1238		0	. 1866		0.2	2416		
2 011111 011213 011301	2	0.	1444		0	. 4245		0.4	1504		

0.2430					
3 0.2575	0.2098	0.8	3663	0.6869	
4	0.1374	0.2	2050	0.4000	
0.1625					
worst s 0 1 2 3 4	ymmetry wors 0.4601 0.2750 0.3613 0.6638 0.2364	st fractal di	imension ta 0.11890 0.08902 0.08758 0.17300 0.07678	rget 0.0 0.0 0.0 0.0 0.0	
[5 rows x	31 columns]				
df_cancer.	tail()				
mean smoothness		texture mea	an perimeter	mean area	mean
564	21.56	22.39	142.00	1479.0	
0.11100 565 0.09780	20.13	28.25	131.20	1261.0	
566 0.08455	16.60	28.08	108.30	858.1	
567 0.11780	20.60	29.33	140.10	1265.0	
568 0.05263	7.76	24.54	47.92	181.0	
	compactness	mean concavi	ity mean co	ncave points	mean
symmetry 564	0.11590	0.243	390	0.13890	
0.1726 565	0.10340	0.144	100	0.09791	
0.1752 566	0.10230	0.092	251	0.05302	
0.1590 567	0.27700	0.351	140	0.15200	
0.2397 568 0.1587	0.04362	0.000	000	0.00000	
mean	fractal dime	nsion v	vorst textur	e worst per	imeter
worst area	\	95623	26.4	•	166.10
2027.0 565		95533	38.2		155.00
1731.0 566		95648	34.1		126.70

1124.0				
567	0.07016	3	39.42	184.60
1821.0		_		
568	0.05884	3	30.37	59.16
268.6				
			vorst concavity	
564 565	0.14100 0.11660	0.21130 0.19220	0.4107 0.3215	
566	0.11390	0.30940	0.3403	
567	0.16500	0.86810	0.9387	
568	0.08996	0.06444	0.0000	
	concave points w	orst symmetry	worst fractal	dimension
target 564	0.2216	0.2060		0.07115
0.0	0.2210	0.2000		0.0/115
565	0.1628	0.2572		0.06637
0.0	0.1020	0.2372		0.00037
566	0.1418	0.2218		0.07820
0.0	0.1410	0.2210		0.07020
567	0.2650	0.4087		0.12400
0.0	012030	011007		0112100
568	0.0000	0.2871		0.07039
1.0				
[5 rows x 3]	_			
acacoa.	330. 230 ()			
count 569 mean 14 std 3	radius mean tex .000000 569.00 .127292 19.28 .524049 4.30	0000 5 ⁶ 9.0 9649 91.9 1036 24.2	meter mean 000000 569.00 069033 654.88 98981 351.91 790000 143.50	0000 9104 4129
	.981000 9.71 .700000 16.17		.70000 420.30	
	370000 18.84		40000 420.30	
	780000 21.80		.00000 782.70	
	110000 39.28		00000 702.70 00000 2501.00	
max 20	33120	10013	2501100	
mean	smoothness mean	compactness m	nean 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	0.06	4920	0.0295	60	
0.020310 50%	0.095870	0.09	2630	0.0615	40	
0.033500	0 105200	0 12	0400	0 1207	00	
75% 0.074000	0.105300	0.13	0400	0.1307	טט	
max 0.201200	0.163400	0.34	5400	0.4268	90	
	symmetry m 69.000000	ean fractal di 560	imension 9.000000		st texture 569.000000	\
mean	0.181162		0.062798		25.677223	
std min	0.027414 0.106000		0.007060 0.049960		6.146258 12.020000	
25%	0.161900		0.057700		21.080000	
50% 75%	0.179200 0.195700		0.061540 0.066120		25.410000 29.720000	
max	0.304000		0.097440		49.540000	
wors	t perimeter	worst area	worst sm	noothness	worst	
compactness	Ì		F.0	0.00000		
count 569.000000	569.000000	569.000000	50	9.000000		
mean 0.254265	107.261213	880.583128		0.132369		
std 0.157336	33.602542	569.356993		0.022832		
min 0.027290	50.410000	185.200000		0.071170		
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	231120000	.25 .100000		0.222000		
wors	t concavity	worst concav	e points	worst sy	mmetry \	
count	569.000000 0.272188		9.000000 0.114606		000000 290076	
mean std	0.208624		0.065732		290070 061867	
min 25%	0.000000 0.114500		0.000000 0.064930		156500 250400	
50%	0.226700		0.004930		282200	
75% max	0.382900 1.252000		0.161400 0.291000		317900 663800	
				0.	00000	
wors count	t fractal di 569		target 000000			

```
0.083946
                                  0.627417
mean
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
                      0.207500
                                  1.000000
max
[8 rows x 31 columns]
df cancer.shape
(569, 31)
```

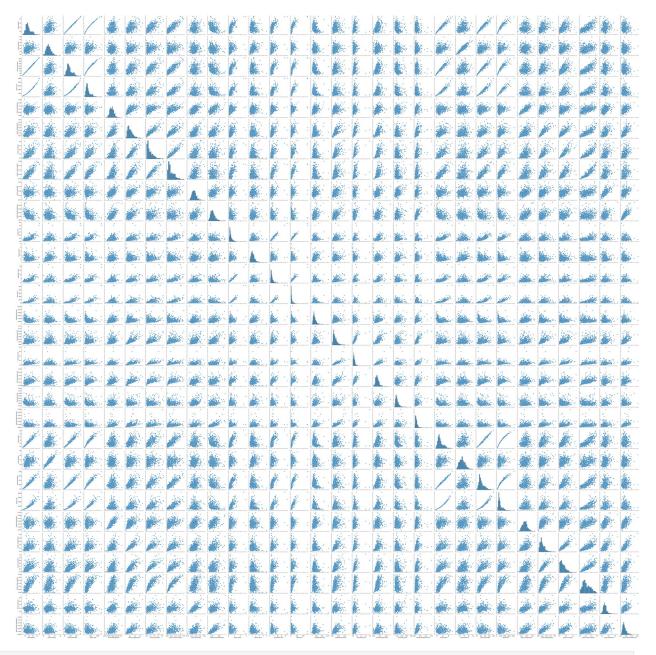
Performing the Data Cleaning

```
df cancer.duplicated().sum()
0
df_cancer.isnull().sum()
mean radius
                            0
                            0
mean texture
                            0
mean perimeter
                            0
mean area
mean smoothness
                            0
                            0
mean compactness
                            0
mean concavity
                            0
mean concave points
                            0
mean symmetry
mean fractal dimension
                            0
radius error
                            0
texture error
                            0
perimeter error
                            0
area error
                            0
smoothness error
                            0
compactness error
concavity error
                            0
concave points error
                            0
symmetry error
fractal dimension error
                            0
worst radius
                            0
worst texture
                            0
worst perimeter
                            0
                            0
worst area
                            0
worst smoothness
worst compactness
                            0
worst concavity
                            0
worst concave points
                            0
```

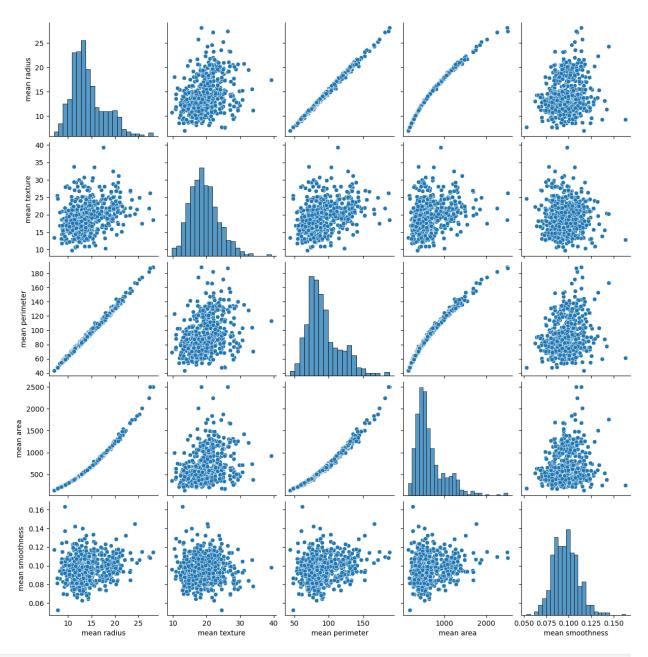
```
worst symmetry 0
worst fractal dimension 0
target 0
dtype: int64
```

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'])
```

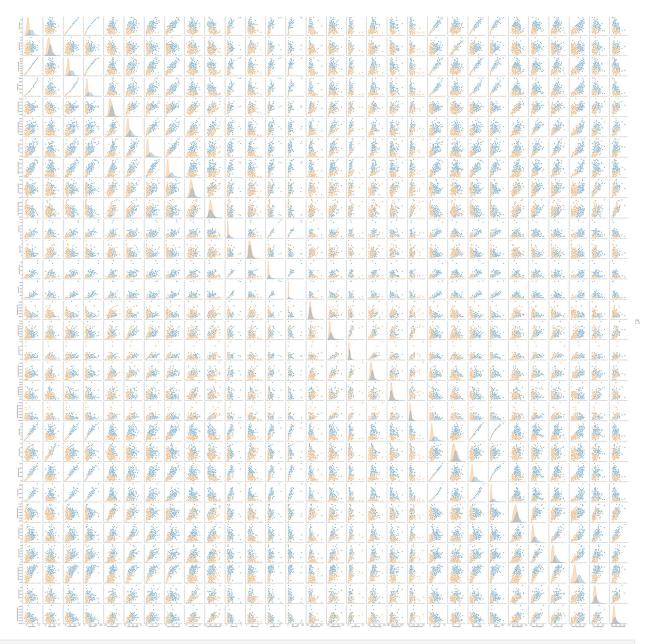


#taking only 5 variable out of 30 just to showcase how powerfull
seaborn library actually is
sns.pairplot(df_cancer, vars = ['mean radius', 'mean texture', 'mean
perimeter', 'mean area',
 'mean smoothness'])
<seaborn.axisgrid.PairGrid at 0x1557612e490>



```
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>
```



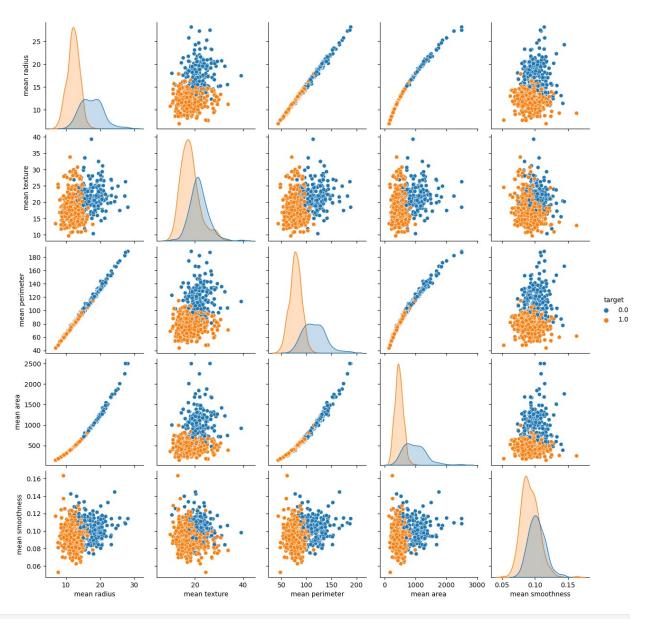
#in above plotting we are not able to differentiate much, so we use
'hue' on target column, which will seperate the two(Malignent,
Benign).
sns.pairplot(df_cancer, hue = 'target', vars = ['mean radius', 'mean

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.

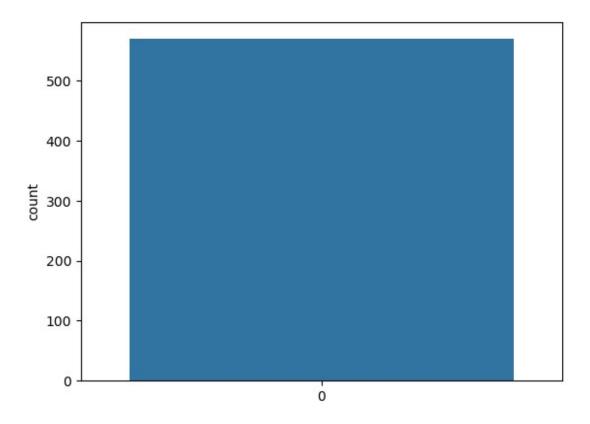
#orange points are not very severe or life threatning

<seaborn.axisgrid.PairGrid at 0x1557cfe4d90>



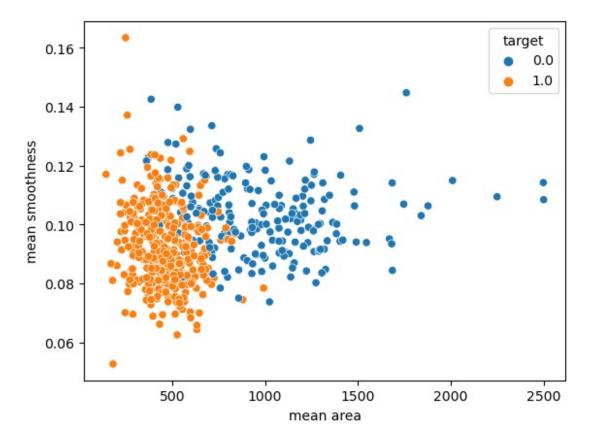
#will simply tell you how many Malignent and Benign cases we have. $\#Malignent = 200 \sim and Benign = 350 \sim approx.$ sns.countplot(df_cancer['target'])

<Axes: ylabel='count'>

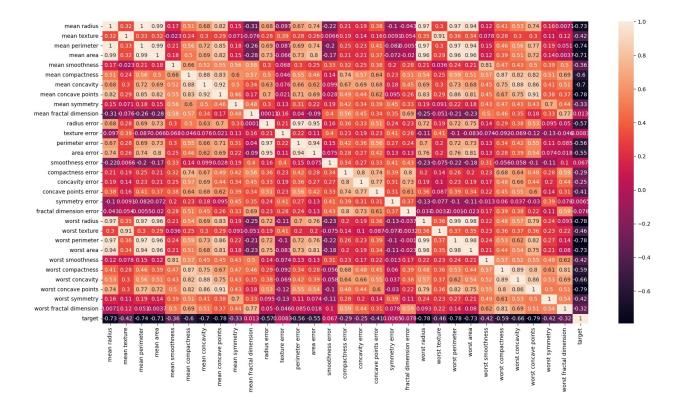


#plotting a scatter plot diagram for mean area anf mean smoothness,
you can plot any feature combination scatterplot.
sns.scatterplot(x='mean area',y='mean smoothness',hue='target',data
=df_cancer)

<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)
```



Splitting the Dataset

```
x = df cancer.drop(['target'],axis =1)
#train test split
Х
     mean radius mean texture mean perimeter mean area
                                                               mean
smoothness
            17.99
                           10.38
                                           122.80
                                                       1001.0
0.11840
                           17.77
           20.57
                                           132.90
                                                       1326.0
0.08474
                                                       1203.0
                           21.25
            19.69
                                           130.00
0.10960
            11.42
                           20.38
                                            77.58
                                                        386.1
0.14250
            20.29
                           14.34
                                                       1297.0
                                           135.10
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
```

F.C.7	20.00	20. 22		140 10	1265 0	
567 0.11780	20.60	29.33		140.10	1265.0	
568	7.76	24.54		47.92	181.0	
0.05263	7170	21131		17132	10110	
	compactness	mean conca	vity n	nean concav	e points n	nean
symmetry 0	0.27760	0.30	9010		0.14710	
0.2419	0.27700	0.30	9010		0.14/10	
1	0.07864	0.0	3690		0.07017	
0.1812						
2	0.15990	0.19	9740		0.12790	
0.2069	0 20200	0.2	4140		0 10520	
3 0.2597	0.28390	0.24	4140		0.10520	
4	0.13280	0.19	9800		0.10430	
0.1809						
 E64	0 11500	0.2	1200		0 12000	
564 0.1726	0.11590	0.24	4390		0.13890	
565	0.10340	0.14	4400		0.09791	
0.1752						
566	0.10230	0.09	9251		0.05302	
0.1590	0 27700	0.21	=140		0 15200	
567 0.2397	0.27700	0.33	5140		0.15200	
568	0.04362	0.00	9000		0.00000	
0.1587						
	6 . 7 . 1 .					,
mean 0	fractal dime	ension 07871	worst	radius wo 25.380	rst texture 17.33	
1		05667		24.990	23.41	
2		05999		23.570	25.53	
3	0.	09744		14.910	26.50	
4	0.	05883		22.540	16.67	7
 EG4	0	05633		25 450	26.40	
564 565		05623 05533		25.450 23.690	26.46 38.25	
566		05648		18.980	34.12	
567		07016		25.740	39.42	
568		05884		9.456	30.37	
wors	t perimeter	worst area	worst	smootnness	worst con	npactness
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
_	132 : 30	170310		3.17770		01 12 730

```
3
                98.87
                             567.7
                                               0.20980
                                                                    0.86630
               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
               0.7119
                                       0.2654
                                                         0.4601
1
                                       0.1860
                                                         0.2750
               0.2416
2
                                       0.2430
               0.4504
                                                         0.3613
3
               0.6869
                                       0.2575
                                                         0.6638
4
                                                         0.2364
               0.4000
                                       0.1625
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
                                                         0.2871
568
               0.0000
                                       0.0000
     worst fractal dimension
0
                       0.11890
1
                       0.08902
2
                       0.08758
3
                       0.17300
4
                       0.07678
                       0.07115
564
565
                       0.06637
                       0.07820
566
567
                       0.12400
568
                       0.07039
[569 rows x 30 columns]
y= df_cancer['target']
У
```

```
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 \overline{d}ata. 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
                          17.57
                                           72.49
                                                       399.8
410
          11.360
0.08858
          18.080
                          21.84
                                          117.40
197
                                                      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
. .
. . .
          13.000
                          21.82
                                           87.50
8
                                                       519.8
0.12730
          13.800
                          15.79
                                           90.43
                                                       584.1
0.10070
                                                       994.0
          17.910
                          21.02
                                          124.40
400
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 me	an concavity	mean concave points	mean
symmetry 306	0.05251	0.001461	0.003261	
0.1632	0.00=0=	0.002.02	0.000_0_	
410	0.05313	0.027830	0.021000	
0.1601 197	0.08642	0.110300	0.057780	
0.1770	0.16600	0. 220000	0.050410	
376 0.2188	0.16600	0.228000	0.059410	
244	0.15580	0.204900	0.088860	
0.1978				
8	0.19320	0.185900	0.093530	
0.2350	0.13320	0.20000	0.033330	
73	0.12800	0.077890	0.050690	
0.1662 400	0.25760	0.318900	0.119800	
0.2113	0.23700	0.518900	0.119000	
118	0.17520	0.213300	0.094790	
0.2096	0 07222	0.017560	0.010520	
206 0.1934	0.07232	0.017560	0.019520	
0.1551				
	fractal dimensi		radius worst text	•
306 410	0.058 0.059			.45
197	0.053			.70
376	0.084			.82
244	0.060	00	21.65 30	.53
8	0.073	 80	15.49 30	.73
73	0.065			.86
400	0.071			.78
118	0.073			.50
206	0.062	85	10.42 23	.22
worst	t perimeter wor	st 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

		106 20	720.2		0 17020	0 5401
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
306 410 197 376 244 8 73 400 118 206	worst	concavity 0.01120 0.18110 0.25350 0.60300 0.34580 0.53900 0.27790 0.90340 0.73560 0.06213	worst concave	e points 0.02500 0.08698 0.09181 0.14650 0.15640 0.20600 0.13830 0.19640 0.20340 0.05588	worst symmet 0.26 0.29 0.23 0.25 0.29 0.43 0.25 0.32 0.32	51 73 69 97 20 78 89 45
306 410 197 376 244 8 73 400 118 206	worst	fractal di	mension 0.08385 0.07745 0.06558 0.12000 0.07614 0.10720 0.10300 0.11980 0.12520 0.07380			
[455	rows x	30 column	ns]			
y_tra	ain					
306 410 197 376 244	1.0 1.0 0.0 1.0 0.0					

73 0.0 400 0.0 118 0.0 206 1.0	9 9 9	455, dtype: f	1 oa+64		
Ivallie: car	get, Length.	455, acype: 1	100104		
x_test					
mean smoothnes		texture mea	n perimeter	mean area	mean
28 0.10820	15.30	25.27	102.40	732.4	
163 0.10120	12.34	22.22	79.85	464.5	
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 0.08320	15.13	29.81	96.71	719.5	
515 0.10490	11.34	18.61	72.76	391.2	
186	18.31	18.58	118.60	1041.0	
0.08588	11.42	20.38	77.58	386.1	
0.14250 261	17.35	23.06	111.00	933.1	
0.08662					
mean symmetry		mean concavi	ty mean con	cave points	mean
28 0.1926	0.16970	0.168	30	0.08751	
163	0.10150	0.053	70	0.02822	
0.1551 123	0.10990	0.088	42	0.05778	
0.1856 361	0.06373	0.033	44	0.02424	
0.1815 549	0.06602	0.015	48	0.00816	
0.1976					
414 0.1852	0.04605	0.046	86	0.02739	
515	0.08499	0.043	92	0.02594	

0.19	27						
186		0.08468	0.0	8169		0.05814	
0.16 3	21	0.28390	e 2	4140		0.10520	
0.25	97	0.20550	0.2	4140		0.10520	
261 0.15	64	0.06290	0.0	2891		0.02837	
0.13	04						
28	mean	fractal dim	005.40	worst r	adius wor 20.27	rst texture 36.71	\
163			.06540		13.58	28.68	
123			.06402		15.70	15.98	
361 549			.05696 .06328		14.20 13.03	29.20 31.45	
414 515			.05294 .06211		17.26 12.47	36.91 23.03	
186			.05425		21.31	26.36	
3			.09744		14.91	26.50	
261		Θ	.05307		19.85	31.47	
	worst	perimeter	worst area	worst s	smoothness	worst comp	pactness
\ 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		120 20					0 14960
201		128.20	1218.0		0.1240		0.14860
	vorst	concovity	worst sons	wa naint	s vorst o	numotov \	
28	WUISL	concavity 0.63350	worst conca	0.2024		symmetry \ 0.4027	
163		0.16880		0.0819)4	0.2268	
123 361		0.25600 0.12120		0.1221 0.0561		0.2889 0.2637	
331		0.12120		0.0501		312037	

```
549
             0.06194
                                     0.03264
                                                        0.3059
414
              0.15470
                                     0.06575
                                                        0.3233
515
             0.16240
                                     0.08542
                                                        0.3060
                                                        0.3206
186
             0.35380
                                     0.15710
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

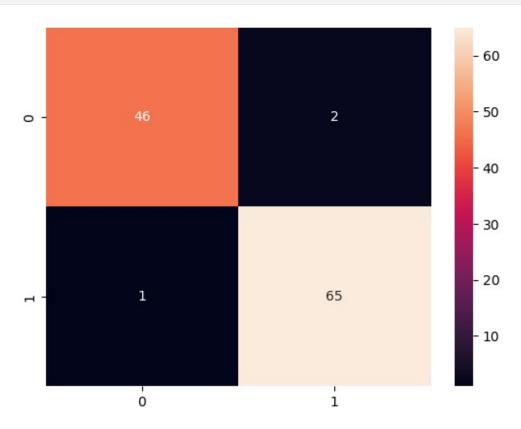
from sklearn.metrics import classification_report, confusion_matrix

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

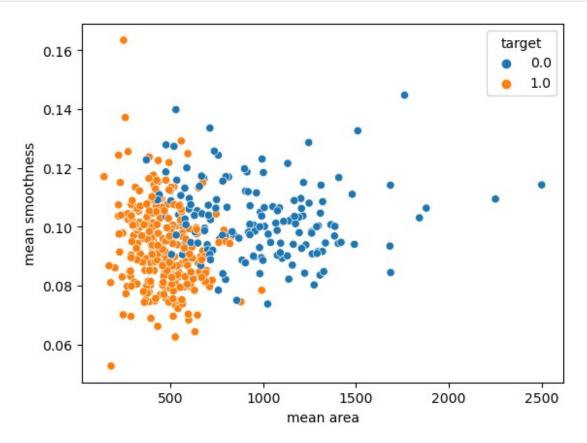
```
y_predict =svm_model.predict(x_test)
y_predict
array([0., 1., 1., 1., 1., 0., 1., 1., 1., 1., 1., 1., 0., 1., 1., 1.,
       1., 1., 1., 0., 1., 1., 1., 1., 0., 0., 1., 0., 0., 0., 1.,
0.,
      1., 1., 0., 1., 1., 0., 1., 1., 0., 1., 1., 0., 0., 1., 0.,
1.,
      1., 1., 1., 1., 0., 0., 0., 1., 0., 0., 0., 1., 1., 1., 1.,
1.,
      1., 0., 1., 0., 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.,
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: >
```



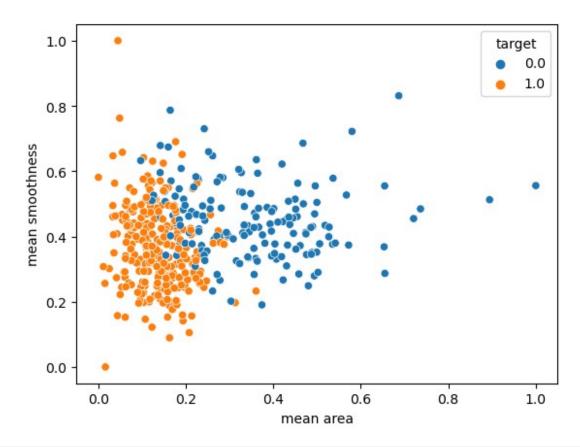
Model Improvisation

```
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: >
```

