BREAST CANCER PROGNOSIS

Determine the Recurrence time for Recurrent type of Cancer using Machine Learning in Python.

Overview

Build a classifier to predict the type of Cancer as Recurrent or Non-recurrent with high accuracy using Random Forest Classifier.And made a regression model to predict the recurrence time for Recurrent type of Cancer using Polynomial Regression of order 2.The model uses machine learning in python language. The model will be providing insights to the doctors regarding time after which cancer can reoccur.

DataSet-

1) ID
2) Outcome (R = recur, N = nonrecur)
3) Time (recurrence time if field 2 = R, disease-free time if field 2 = N)
Ten real-valued features are computed for each cell nucleus (4-33)-:
    a) radius (mean of distances from center to points on the perimeter)
    b) texture (standard deviation of gray-scale values)
    c) perimeter
    d) area
    e) smoothness (local variation in radius lengths)
    f) compactness (perimeter^2 / area - 1.0)
    g) concavity (severity of concave portions of the contour)
    h) concave points (number of concave portions of the contour)
    i) symmetry
    j) fractal dimension ("coastline approximation" - 1)

4)Lymph Node Status

Approach-

In this Project various Libraries of Python like-Pandas ,Numpy,Matplotlib,seaborn,scikit learn ,imblearn were used.Apart from this SMOTE technique was used to handle imbalance nature of the dataset by oversampling of minority class.