CA707 Data Mining Lab

Project Report- Nov 2019

Roll No: 205118033

Name: Kushbinder singh

Title of the Project :

Breast Cancer Tumour Classification (Malignant or Benign) using Knn Classification

Dataset Description:

Name : breast-cancer-Wisconsin

URL : https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+(Diagnostic)

No of Records: 699

No of Features: 11

List all Features(With One Line Description):

One of the old methods of diagnosis of Breast Cancer is carried out by Fine Needle Aspiration procedure. A thin and hollow needle is inserted into a lump and a sample of cells is extracted. The specimen is then studied under a microscope and the following measurements are made.

Feature	Description
Id	Patient Sample code number
clump_thickness	Determines if the cells are mono-layered or multi-layered
unif_cell_size	A measure of the variance of the sizes of the cells.
unif_cell_shape	A measure of the variance in the shapes of the cells. Cancer cells vary in shape.
marg_adhesion	Assess the adhesion ability of the cells. Cancer cells tend not to stick together.

single_epith_cell_size	Determines if the epithelial cells have signifcantly enlarged.
bare_nuclei	A measurement of the proportion of the number of cells not covered by cytoplasm to those that are.
bland_chrom	A rating of the texture of nucleus from fine to coarse.
norm_nucleoli	Nucleoli are small structures present in the nucleus. A higher value of this attribute would indicate a higher chance of malig-nancy.
mitoses	Describes the level of cell division.
class	(2 for benign, 4 for malignant)

Type of problem: Classification

Algorithm Used: K- nearest neighbours(KNN) Algorithm

Evaluation Metric: Out of 699, 678 predicted correctly and 21 predicted wrong

Accuracy Obtained:

Using Libraries: 97.714%

Your Algorithm: 97.142 %

Date **Signature**