

3.1 Dataset

Source: Breast Cancer General Dataset from Kaggle

Creator: Samples from Dr.Wolberg's Clinical Cases

Method: Using Fluid Samples, Taken From Patients

Tech: Xcvt/Digital Scan, an Easy-To-Use Graphical Computer Program

Develop: Dataset Reflects Chronological Grouping of data

It is a balanced dataset with 16 missing values in column 6 (Single Epithelial

Cell Size)




3.1 Dataset

Overview



The image shows the Kaggle dataset page for the 'Breast Cancer Wisconsin (Diagnostic) Data Set'. The page features a header with the Kaggle logo and a navigation bar. The main content area includes a title, a description, and a 'Data' tab. The description states: 'Predict whether the cancer is benign or malignant'. Below the description, there are links for 'Data', 'Kernels (1983)', 'Discussion (23)', 'Activity', and 'Metadata'. A 'Download (122 KB)' button and a 'New Notebook' button are also visible. The page is updated 3 years ago (Version 2).



The image shows a summary and data preview of the Breast Cancer Wisconsin (Diagnostic) Data Set. The summary includes histograms for the 'diagnosis' variable (B for benign, M for malignant) and the 'radius_mean', 'texture_mean', 'perimeter_mean', and 'area' variables. The data preview shows a table with columns: id, diagnosis, radius_mean, texture_mean, perimeter_mean, and area.

id	diagnosis	radius_mean	texture_mean	perimeter_mean	area
1	B	17.99	16.34	123.6	
2	B	28.57	17.77	150.9	
3	B	19.69	21.35	138	
4	B	11.42	20.38	77.58	
5	B	38.29	14.34	125.1	
6	B	12.45	15.7	82.57	
7	B	18.25	16.98	119.6	
8	B	13.71	20.83	96.3	
9	B	15	21.02	87.5	
10	B	12.45	24.64	80.97	
11	B	16.82	23.24	162.7	
12	B	15.78	17.89	183.6	
13	B	19.17	24.8	133.4	
14	B	15.85	23.46	183.7	
15	B	13.72	23.61	93.6	
16	B	14.54	27.54	96.73	
17	B	14.68	26.13	64.74	
18	B	15.15	26.68	186.1	
19	B	19.81	23.15	150	
20	B	13.54	14.36	87.46	
21	B	13.88	16.71	85.63	
22	B	9.586	12.44	69.34	