Breast Cancer Wisconsin (Original)

Donated on 7/14/1992

Original Wisconsin Breast Cancer Database

Dataset Characteristics
Multivariate

Subject Area
Health and Medicine

Associated Tasks
Classification

Feature Type
Integer

Instances
699

Features
9

Dataset Information

Additional Information

Samples arrive periodically as Dr. Wolberg reports his clinical cases. The database therefore reflects this chronological grouping of the data. This grouping information appears immediately below, having been removed from the data itself:

Group 1: 367 instances (January 1989)
Group 2: 70 instances (October 1989)
Group 3: 31 instances (February 1990)
Group 4: 17 instances (April 1990)
Group 5: 48 instances (August 1990)
Group 6: 49 instances (Updated January 1991)
Group 7: 31 instances (June 1991)
Group 8: 86 instances (November 1991)

Total: 699 points (as of the donated datbase on 15 July 1992)

Note that the results summarized above in Past Usage refer to a dataset of size 369, while Group 1 has only 367 instances. This is because it originally contained 369 instances; 2 were

removed. The following statements summarizes changes to the original Group 1's set of data:

Group 1 : 367 points: 200B 167M (January 1989)

Revised Jan 10, 1991: Replaced zero bare nuclei in 1080185 & 1187805

Revised Nov 22,1991: Removed 765878,4,5,9,7,10,10,10,3,8,1 no record
: Removed 484201,2,7,8,8,4,3,10,3,4,1 zero epithelial
: Changed 0 to 1 in field 6 of sample 1219406
: Changed 0 to 1 in field 8 of following sample:
: 1182404,2,3,1,1,1,2,0,1,1,1

SHOW LESS

Has Missing Values?
Yes

Variables Table

Variable Name	Role	Type	Description	Units	Missing Values
Sample_code_number	ID	Categorical			no
Clump_thickness	Feature	Integer			no
Uniformity_of_cell_size	Feature	Integer			no
Uniformity_of_cell_shape	Feature	Integer			no
Marginal_adhesion	Feature	Integer			no
Single_epithelial_cell_size	Feature	Integer			no
Bare_nuclei	Feature	Integer			yes
Bland_chromatin	Feature	Integer			no
Normal_nucleoli	Feature	Integer			no
Mitoses	Feature	Integer			no

Rows per page 10 0 to 10 of 11

Additional Variable Information

1. Sample code number: id number
2. Clump Thickness: 1 - 10
3. Uniformity of Cell Size: 1 - 10
4. Uniformity of Cell Shape: 1 - 10
5. Marginal Adhesion: 1 - 10
6. Single Epithelial Cell Size: 1 - 10
7. Bare Nuclei: 1 - 10
8. Bland Chromatin: 1 - 10
9. Normal Nucleoli: 1 - 10
10. Mitoses: 1 - 10
11. Class: (2 for benign, 4 for malignant)

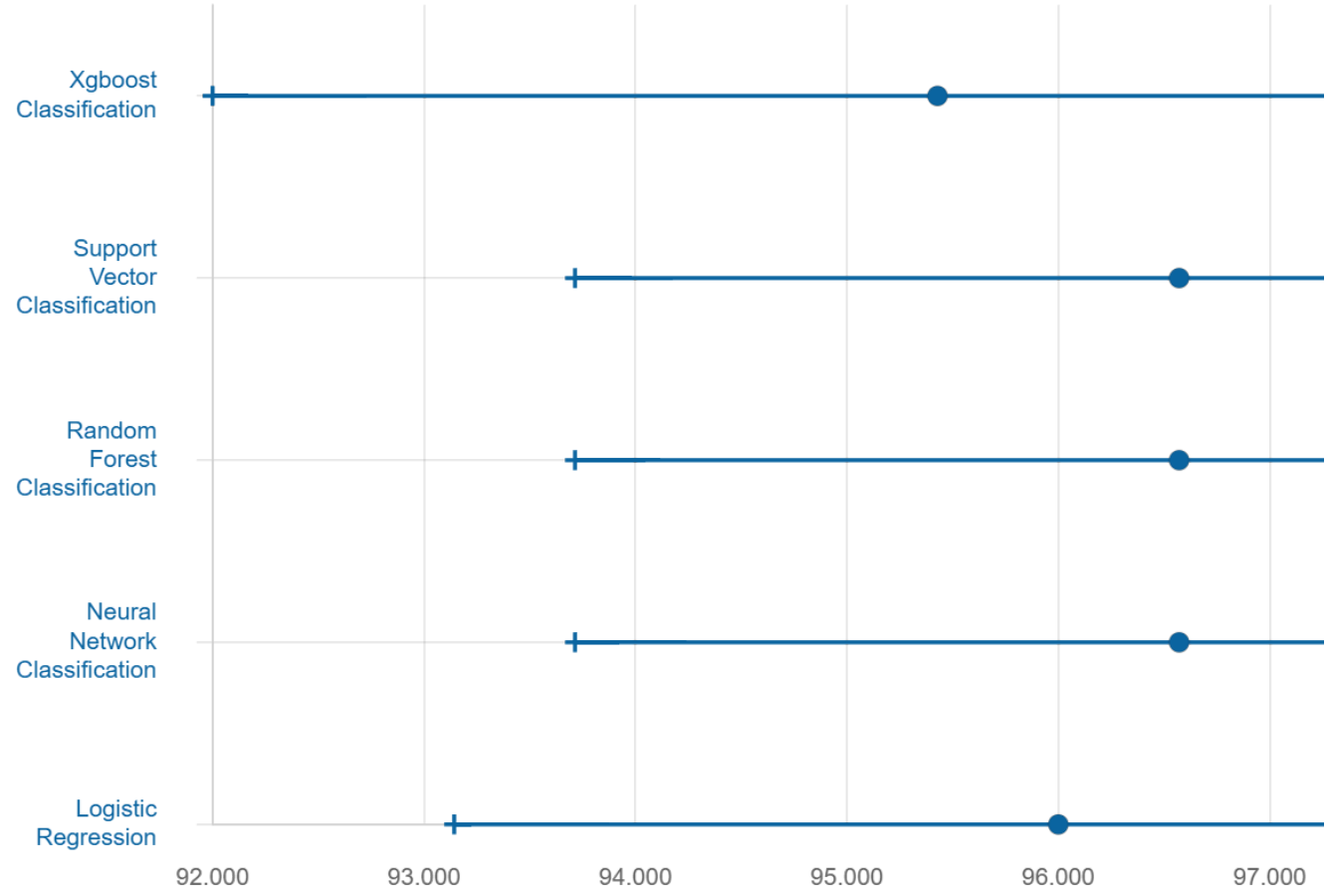
SHOW LESS

Class Labels

2 = benign, 4 = malignant

Baseline Model Performance

Accuracy | Precision



Dataset Files

File	Size
wdbc.data	121.2 KB
wpbc.data	43.2 KB
unformatted-data	20.9 KB
breast-cancer-wisconsin.data	19.4 KB
wpbc.names	5.5 KB

Rows per page 5 0 to 5 of 8

Papers Citing this Dataset

SORT BY YEAR, DESC

On the Bias of Precision Estimation Under Separate Sampling
By Shuilian Xie, Ulisses Braga-Neto. 2019
Published in Cancer informatics.

Rbf Kernel Optimization Method with Particle Swarm Optimization on Svm Using the Analy...
By Rarasmaya Indraswari, Agus Arifin. 2017
Published in Jurnal Ilmu Komputer dan Informasi.

Iteratively Reweighted Least Squares Algorithms for L1-Norm Principal Component Analysis
By Young Park, Diego Klabjan. 2016
Published in ArXiv.

Widened KRIMP: Better Performance through Diverse Parallelism
By Oliver Sampson, Michael Berthold. 2014
Published in IDA.

Automated Empirical Selection of Rule Induction Methods Based on Recursive Iteration of R...
By Shusaku Tsumoto, Shoji Hirano. 2010
Published in 2010 IEEE International Conference on Data Mining Workshops.

Rows per page 5 0 to 5 of 6

Reviews

There are no reviews for this dataset yet.

WRITE A REVIEW

DOWNLOAD (84.5 KB)

IMPORT IN PYTHON

CITE

Citations/Acknowledgements

If you use this dataset, please cite:

This breast cancer databases was obtained from the University of Wisconsin Hospitals, Madison from Dr. William H. Wolberg. If you publish results when using this database, then please include this information in your acknowledgements. Also, please cite one or more of:
1. O. L. Mangasarian and W. H. Wolberg: "Cancer diagnosis via linear programming", SIAM News, Volume 23, Number 5, September 1990, pp 1 & 18....

Keywords

health cancer

Creators

WIlliam Wolberg

DOI

10.24432/C5HP4Z

License

About Us

CML

National Science Foundation

NAVIGATION

Home

View Datasets

Donate a Dataset

LOGISTICS

Contact

Privacy Notice

Feature Request or Bug Report