Haberman's Survival Data Set

Download: Data Folder, Data Set Description

Abstract: Dataset contains cases from study conducted on the survival of patients who had undergone surgery for breast cancer

Data Set Characteristics:	Multivariate	**Number of Instances:**	306	**Area:**	Life
Attribute Characteristics:	Integer	**Number of Attributes:**	3	**Date Donated**	1999-03-04
Associated Tasks:	Classification	**Missing Values?**	No	**Number of Web Hits:**	76398

Source:

Donor:

Tjen-Sien Lim (limt '@' stat.wisc.edu)

Data Set Information:

The dataset contains cases from a study that was conducted between 1958 and 1970 at the University of Chicago's Billings Hospital on the survival of patients who had undergone surgery for breast cancer.

Attribute Information:

1. Age of patient at time of operation (numerical)
2. Patient's year of operation (year - 1900, numerical)
3. Number of positive axillary nodes detected (numerical)
4. Survival status (class attribute)
-- 1 = the patient survived 5 years or longer
-- 2 = the patient died within 5 year

Relevant Papers:

Haberman, S. J. (1976). Generalized Residuals for Log-Linear Models, Proceedings of the 9th International Biometrics Conference, Boston, pp. 104-122.

Landwehr, J. M., Pregibon, D., and Shoemaker, A. C. (1984), Graphical Models for Assessing Logistic Regression Models (with discussion), Journal of the American Statistical Association 79: 61-83.
[Web Link]

Lo, W.-D. (1993). Logistic Regression Trees, PhD thesis, Department of Statistics, University of Wisconsin, Madison, WI.
[Web Link]

Papers That Cite This Data Set[1]:

Dennis DeCoste. Anytime Query-Tuned Kernel Machines via Cholesky Factorization. SDM. 2003. [View Context].

Dennis DeCoste. Anytime Interval-Valued Outputs for Kernel Machines: Fast Support Vector Machine Classification via Distance Geometry. ICML. 2002. [View Context].

Yin Zhang and W. Nick Street. Bagging with Adaptive Costs. Management Sciences Department University of Iowa Iowa City. [View Context].

Denver Dash and Gregory F. Cooper. Model Averaging with Discrete Bayesian Network Classifiers. Decision Systems Laboratory Intelligent Systems Program University of Pittsburgh. [View Context].

Citation Request:

Please refer to the Machine Learning Repository's citation policy

---

[1] Papers were automatically harvested and associated with this data set, in collaboration with Rexa.info

Supported By: 

In Collaboration With: