Repository Web

Search

View ALL Data Sets

Center for Machine Learning and Intelligent Systems

Mushroom Data Set

Download: Data Folder, Data Set Description

Abstract: From Audobon Society Field Guide; mushrooms described in terms of physical characteristics; classification: poisonous or edible

Data Set Characteristics:	Multivariate	Number of Instances:	8124	**Area:**	Life
Attribute Characteristics:	Categorical	**Number of Attributes:**	22	**Date Donated**	1987-04-27
Associated Tasks:	Classification	**Missing Values?**	Yes	**Number of Web Hits:**	238994

Source:

Origin:

Mushroom records drawn from The Audubon Society Field Guide to North American Mushrooms (1981). G. H. Lincoff (Pres.), New York: Alfred A. Knopf

Donor:

Jeff Schlimmer (Jeffrey.Schlimmer '@' a.gp.cs.cmu.edu)

Data Set Information:

This data set includes descriptions of hypothetical samples corresponding to 23 species of gilled mushrooms in the Agaricus and Lepiota Family (pp. 500-525). Each species is identified as definitely edible, definitely

poisonous, or of unknown edibility and not recommended. This latter class was combined with the poisonous one. The Guide clearly states that there is no simple rule for determining the edibility of a mushroom; no rule like ``leaflets three, let it be'' for Poisonous Oak and Ivy.

Attribute Information:

1. cap-shape: bell=b,conical=c,convex=x,flat=f, knobbed=k,sunken=s
2. cap-surface: fibrous=f,grooves=g,scaly=y,smooth=s
3. cap-color: brown=n,buff=b,cinnamon=c,gray=g,green=r, pink=p,purple=u,red=e,white=w,yellow=y
4. bruises?: bruises=t,no=f
5. odor: almond=a,anise=l,creosote=c,fishy=y,foul=f, musty=m,none=n,pungent=p,spicy=s
6. gill-attachment: attached=a,descending=d,free=f,notched=n
7. gill-spacing: close=c,crowded=w,distant=d
8. gill-size: broad=b,narrow=n
9. gill-color: black=k,brown=n,buff=b,chocolate=h,gray=g, green=r,orange=o,pink=p,purple=u,red=e, white=w,yellow=y
10. stalk-shape: enlarging=e,tapering=t
11. stalk-root: bulbous=b,club=c,cup=u,equal=e, rhizomorphs=z,rooted=r,missing=?
12. stalk-surface-above-ring: fibrous=f,scaly=y,silky=k,smooth=s
13. stalk-surface-below-ring: fibrous=f,scaly=y,silky=k,smooth=s
14. stalk-color-above-ring: brown=n,buff=b,cinnamon=c,gray=g,orange=o, pink=p,red=e,white=w,yellow=y
15. stalk-color-below-ring: brown=n,buff=b,cinnamon=c,gray=g,orange=o, pink=p,red=e,white=w,yellow=y
16. veil-type: partial=p,universal=u
17. veil-color: brown=n,orange=o,white=w,yellow=y
18. ring-number: none=n,one=o,two=t
19. ring-type: cobwebby=c,evanescent=e,flaring=f,large=l, none=n,pendant=p,sheathing=s,zone=z
20. spore-print-color: black=k,brown=n,buff=b,chocolate=h,green=r, orange=o,purple=u,white=w,yellow=y
21. population: abundant=a,clustered=c,numerous=n, scattered=s,several=v,solitary=y
22. habitat: grasses=g,leaves=l,meadows=m,paths=p, urban=u,waste=w,woods=d

Relevant Papers:

Schlimmer,J.S. (1987). Concept Acquisition Through Representational Adjustment (Technical Report 87-19). Doctoral disseration, Department of Information and Computer Science, University of California, Irvine.
[Web Link]

Iba,W., Wogulis,J., & Langley,P. (1988). Trading off Simplicity and Coverage in Incremental Concept Learning. In Proceedings of the 5th International Conference on Machine Learning, 73-79. Ann Arbor, Michigan: Morgan Kaufmann.
[Web Link]

Duch W, Adamczak R, Grabczewski K (1996) Extraction of logical rules from training data using backpropagation networks, in: Proc. of the The 1st Online Workshop on Soft Computing, 19-30.Aug.1996, pp. 25-30, [Web Link]
[Web Link]

Duch W, Adamczak R, Grabczewski K, Ishikawa M, Ueda H, Extraction of crisp logical rules using constrained backpropagation networks - comparison of two new approaches, in: Proc. of the European Symposium on Artificial Neural Networks (ESANN'97), Bruge, Belgium 16-18.4.1997.
[Web Link]

Papers That Cite This Data Set[1]:

Manuel Oliveira. Library Release Form Name of Author: Stanley Robson de Medeiros Oliveira Title of Thesis: Data Transformation For Privacy-Preserving Data Mining Degree: Doctor of Philosophy Year this Degree Granted. University of Alberta Library. 2005. [View Context].

Hyunsoo Kim and Se Hyun Park. Data Reduction in Support Vector Machines by a Kernelized Ionic Interaction Model. SDM. 2004. [View Context].

Xiaoyong Chai and Li Deng and Qiang Yang and Charles X. Ling. Test-Cost Sensitive Naive Bayes Classification. ICDM. 2004. [View Context].

Daniel J. Lizotte and Omid Madani and Russell Greiner. Budgeted Learning of Naive-Bayes Classifiers. UAI. 2003. [View Context].

Daniel Barbar and Yi Li and Julia Couto. COOLCAT: an entropy-based algorithm for categorical clustering. CIKM. 2002. [View Context].

Stephen D. Bay and Michael J. Pazzani. Detecting Group Differences: Mining Contrast Sets. Data Min. Knowl. Discov, 5. 2001. [View Context].

Jinyan Li and Guozhu Dong and Kotagiri Ramamohanarao and Limsoon Wong. DeEPs: A New Instance-based Discovery and Classification System. Proceedings of the Fourth European Conference on Principles and Practice of Knowledge Discovery in Databases. 2001. [View Context].

Huan Liu and Hongjun Lu and Jie Yao. Toward Multidatabase Mining: Identifying Relevant Databases. IEEE Trans. Knowl. Data Eng, 13. 2001. [View Context].

Jinyan Li and Guozhu Dong and Kotagiri Ramamohanarao. Instance-Based Classification by Emerging Patterns. PKDD. 2000. [View Context].

Farhad Hussain and Huan Liu and Einoshin Suzuki and Hongjun Lu. Exception Rule Mining with a Relative Interestingness Measure. PAKDD. 2000. [View Context].

Kiri Wagstaff and Claire Cardie. Clustering with Instance-level Constraints. ICML. 2000. [View Context].

Mark A. Hall and Lloyd A. Smith. Feature Selection for Machine Learning: Comparing a Correlation-Based Filter Approach to the Wrapper. FLAIRS Conference. 1999. [View Context].

Jinyan Li and Xiuzhen Zhang and Guozhu Dong and Kotagiri Ramamohanarao and Qun Sun. Efficient Mining of High Confidence Association Rules without Support Thresholds. PKDD. 1999. [View Context].

Seth Bullock and Peter M. Todd. Made to Measure: Ecological Rationality in Structured Environments. Center for Adaptive Behavior and Cognition Max Planck Institute for Human Development. 1999. [View Context].

Venkatesh Ganti and Johannes Gehrke and Raghu Ramakrishnan. CACTUS - Clustering Categorical Data Using Summaries. KDD. 1999. [View Context].

Ismail Taha and Joydeep Ghosh. Symbolic Interpretation of Artificial Neural Networks. IEEE Trans. Knowl. Data Eng, 11. 1999. [View Context].

Mark A. Hall. Department of Computer Science Hamilton, NewZealand Correlation-based Feature Selection for Machine Learning. Doctor of Philosophy at The University of Waikato. 1999. [View Context].

Huan Liu and Hongjun Lu and Ling Feng and Farhad Hussain. Efficient Search of Reliable Exceptions. PAKDD. 1999. [View Context].

Huan Liu and Rudy Setiono. Incremental Feature Selection. Appl. Intell, 9. 1998. [View Context].

Nicholas Howe and Claire Cardie. Examining Locally Varying Weights for Nearest Neighbor Algorithms. ICCBR. 1997. [View Context].

Robert M French. Pseudo-recurrent connectionist networks: An approach to the "sensitivity-stability" dilemma.. Connection Science. 1997. [View Context].

Huan Liu and Rudy Setiono. A Probabilistic Approach to Feature Selection - A Filter Solution. ICML. 1996. [View Context].

Kamal Ali and Michael J. Pazzani. Error Reduction through Learning Multiple Descriptions. Machine Learning, 24. 1996. [View Context].

Guszti Bartfai. VICTORIA UNIVERSITY OF WELLINGTON Te Whare Wananga o te Upoko o te Ika a Maui. Department of Computer Science PO Box 600. 1996. [View Context].

Geoffrey I. Webb. OPUS: An Efficient Admissible Algorithm for Unordered Search. J. Artif. Intell. Res. (JAIR, 3. 1995. [View Context].

Wl odzisl/aw Duch and Rafal Adamczak and Krzysztof Grabczewski and Norbert Jankowski. Control and Cybernetics. Department of Computer Methods, Nicholas Copernicus University. [View Context].

Huan Liu. A Family of Efficient Rule Generators. Department of Information Systems and Computer Science National University of Singapore. [View Context].

Shi Zhong and Weiyu Tang and Taghi M. Khoshgoftaar. Boosted Noise Filters for Identifying Mislabeled Data. Department of Computer Science and Engineering Florida Atlantic University. [View Context].

Chotirat Ann and Dimitrios Gunopulos. Scaling up the Naive Bayesian Classifier: Using Decision Trees for

Feature Selection. Computer Science Department University of California. [View Context].

Eric P. Kasten and Philip K. McKinley. MESO: Perceptual Memory to Support Online Learning in Adaptive Software. Proceedings of the Third International Conference on Development and Learning (ICDL. [View Context].

Stefan R uping. A Simple Method For Estimating Conditional Probabilities For SVMs. CS Department, AI Unit Dortmund University. [View Context].

Josep Roure Alcobe. Incremental Hill-Climbing Search Applied to Bayesian Network Structure Learning. Escola Universitria Politcnica de Mataro. [View Context].

Wl odzisl and Rafal Adamczak and Krzysztof Grabczewski and Grzegorz Zal. A hybrid method for extraction of logical rules from data. Department of Computer Methods, Nicholas Copernicus University. [View Context].

Jinyan Li and Kotagiri Ramamohanarao and Guozhu Dong. ICML2000 The Space of Jumping Emerging Patterns and Its Incremental Maintenance Algorithms. Department of Computer Science and Software Engineering, The University of Melbourne, Parkville. [View Context].

Wl/odzisl/aw Duch and Rafal Adamczak and Krzysztof Grabczewski. Extraction of crisp logical rules using constrained backpropagation networks. Department of Computer Methods, Nicholas Copernicus University. [View Context].

Wl odzisl/aw Duch and Rudy Setiono and Jacek M. Zurada. Computational intelligence methods for rule-based data understanding. [View Context].

C. Titus Brown and Harry W. Bullen and Sean P. Kelly and Robert K. Xiao and Steven G. Satterfield and John G. Hagedorn and Judith E. Devaney. Visualization and Data Mining in an 3D Immersive Environment: Summer Project 2003. [View Context].

Daniel J. Lizotte. Library Release Form Name of Author. Budgeted Learning of Naive Bayes Classifiers. [View Context].

David R. Musicant. DATA MINING VIA MATHEMATICAL PROGRAMMING AND MACHINE LEARNING. Doctor of Philosophy (Computer Sciences) UNIVERSITY. [View Context].

Sherrie L. W and Zijian Zheng. A BENCHMARK FOR CLASSIFIER LEARNING. Basser Department of Computer Science The University of Sydney. [View Context].

Anthony Robins and Marcus Frean. Learning and generalisation in a stable network. Computer Science, The University of Otago. [View Context].

Rudy Setiono. Extracting M-of-N Rules from Trained Neural Networks. School of Computing National University of Singapore. [View Context].

Jos'e L. Balc'azar. Rules with Bounded Negations and the Coverage Inference Scheme. Dept. LSI, UPC. [View Context].

Mehmet Dalkilic and Arijit Sengupta. A Logic-theoretic classifier called Circle. School of Informatics Center for

Genomics and BioInformatics Indiana University. [View Context].

Daniel J. Lizotte and Omid Madani and Russell Greiner. Budgeted Learning, Part II: The Na#ve-Bayes Case. Department of Computing Science University of Alberta. [View Context].

Ron Kohavi and Barry G. Becker and Dan Sommerfield. Improving Simple Bayes. Data Mining and Visualization Group Silicon Graphics, Inc. [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:

About || Citation Policy || Donation Policy || Contact || CML