1. Title: Mushroom Database

2. Sources:
   (a) Mushroom records drawn from The Audubon Society Field Guide to North
       American Mushrooms (1981). G. H. Lincoff (Pres.), New York: Alfred
       A. Knopf
   (b) Donor: Jeff Schlimmer (Jeffrey.Schlimmer@a.gp.cs.cmu.edu)
   (c) Date: 27 April 1987

3. Past Usage:
   1. 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.
      --- STAGGER: asymptoted to 95% classification accuracy after reviewing
          1000 instances.
   2. 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.
      -- approximately the same results with their HILLARY algorithm
   3. In the following references a set of rules (given below) were
       learned for this data set which may serve as a point of
       comparison for other researchers.

      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,
      available on-line at: http://www.bioele.nuee.nagoya-u.ac.jp/wsc1/

      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,
      pp. xx-xx

      Wlodzislaw Duch, Department of Computer Methods, Nicholas Copernicus
      University, 87-100 Torun, Grudziadzka 5, Poland
      e-mail: duch@phys.uni.torun.pl
      WWW   http://www.phys.uni.torun.pl/kmk/

      Date: Mon, 17 Feb 1997 13:47:40 +0100
      From: Wlodzislaw Duch <duch@phys.uni.torun.pl>
      Organization: Dept. of Computer Methods, UMK

      I have attached a file containing logical rules for mushrooms.
      It should be helpful for other people since only in the last year I
      have seen about 10 papers analyzing this dataset and obtaining quite
      complex rules. We will try to contribute other results later.

      With best regards, Wlodek Duch
      _____

      Logical rules for the mushroom data sets.

      Logical rules given below seem to be the simplest possible for the
      mushroom dataset and therefore should be treated as benchmark results.

      Disjunctive rules for poisonous mushrooms, from most general
      to most specific:

      P_1) odor=NOT(almond.OR.anise.OR.none)
           120 poisonous cases missed, 98.52% accuracy

      P_2) spore-print-color=green
           48 cases missed, 99.41% accuracy

```
       P_3) odor=none.AND.stalk-surface-below-ring=scaly.AND.
               (stalk-color-above-ring=NOT.brown)
           8 cases missed, 99.90% accuracy


       P_4) habitat=leaves.AND.cap-color=white
               100% accuracy


       Rule P_4) may also be


       P_4') population=clustered.AND.cap_color=white


       These rule involve 6 attributes (out of 22). Rules for edible
       mushrooms are obtained as negation of the rules given above, for
       example the rule:


       odor=(almond.OR.anise.OR.none).AND.spore-print-color=NOT.green


       gives 48 errors, or 99.41% accuracy on the whole dataset.


       Several slightly more complex variations on these rules exist,
       involving other attributes, such as gill_size, gill_spacing,
       stalk_surface_above_ring, but the rules given above are the simplest
       we have found.



4. Relevant 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.


5. Number of Instances: 8124


6. Number of Attributes: 22 (all nominally valued)


7. Attribute Information: (classes: edible=e, poisonous=p)
     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
```

8. Missing Attribute Values: 2480 of them (denoted by "?"), all for
   attribute #11.

9. Class Distribution:
   --   edible: 4208 (51.8%)
   -- poisonous: 3916 (48.2%)
   --   total: 8124 instances