

# Classification Of Mushroom Data

January 19, 2022

## 1 Classification Of Mushroom Data

### Dataset

#### 1.1 Dealing With Dependencies

```
[1]: !pip install -U pandas numpy pandas_profiling[notebook] sklearn catboost  
     ↪seaborn matplotlib ipywidgets
```

```
Requirement already satisfied: pandas in  
/srv/conda/envs/notebook/lib/python3.7/site-packages (1.3.5)  
Requirement already satisfied: numpy in  
/srv/conda/envs/notebook/lib/python3.7/site-packages (1.21.5)  
Requirement already satisfied: pandas_profiling[notebook] in  
/srv/conda/envs/notebook/lib/python3.7/site-packages (3.1.0)  
Requirement already satisfied: sklearn in  
/srv/conda/envs/notebook/lib/python3.7/site-packages (0.0)  
Requirement already satisfied: catboost in  
/srv/conda/envs/notebook/lib/python3.7/site-packages (1.0.4)  
Requirement already satisfied: seaborn in  
/srv/conda/envs/notebook/lib/python3.7/site-packages (0.11.2)  
Requirement already satisfied: matplotlib in  
/srv/conda/envs/notebook/lib/python3.7/site-packages (3.5.1)  
Requirement already satisfied: ipywidgets in  
/srv/conda/envs/notebook/lib/python3.7/site-packages (7.6.5)  
Requirement already satisfied: pytz>=2017.3 in  
/srv/conda/envs/notebook/lib/python3.7/site-packages (from pandas) (2021.3)  
Requirement already satisfied: python-dateutil>=2.7.3 in  
/srv/conda/envs/notebook/lib/python3.7/site-packages (from pandas) (2.8.2)  
Requirement already satisfied: requests>=2.24.0 in  
/srv/conda/envs/notebook/lib/python3.7/site-packages (from  
pandas_profiling[notebook]) (2.26.0)  
Requirement already satisfied: visions[type_image_path]==0.7.4 in  
/srv/conda/envs/notebook/lib/python3.7/site-packages (from  
pandas_profiling[notebook]) (0.7.4)  
Requirement already satisfied: htmlmin>=0.1.12 in  
/srv/conda/envs/notebook/lib/python3.7/site-packages (from  
pandas_profiling[notebook]) (0.1.12)
```

Requirement already satisfied: joblib~=1.0.1 in  
/srv/conda/envs/notebook/lib/python3.7/site-packages (from  
pandas\_profiling[notebook]) (1.0.1)

Requirement already satisfied: missingno>=0.4.2 in  
/srv/conda/envs/notebook/lib/python3.7/site-packages (from  
pandas\_profiling[notebook]) (0.5.0)

Requirement already satisfied: scipy>=1.4.1 in  
/srv/conda/envs/notebook/lib/python3.7/site-packages (from  
pandas\_profiling[notebook]) (1.7.3)

Requirement already satisfied: multimethod>=1.4 in  
/srv/conda/envs/notebook/lib/python3.7/site-packages (from  
pandas\_profiling[notebook]) (1.6)

Requirement already satisfied: jinja2>=2.11.1 in  
/srv/conda/envs/notebook/lib/python3.7/site-packages (from  
pandas\_profiling[notebook]) (3.0.3)

Requirement already satisfied: markupsafe~=2.0.1 in  
/srv/conda/envs/notebook/lib/python3.7/site-packages (from  
pandas\_profiling[notebook]) (2.0.1)

Requirement already satisfied: tqdm>=4.48.2 in  
/srv/conda/envs/notebook/lib/python3.7/site-packages (from  
pandas\_profiling[notebook]) (4.62.3)

Requirement already satisfied: phik>=0.11.1 in  
/srv/conda/envs/notebook/lib/python3.7/site-packages (from  
pandas\_profiling[notebook]) (0.12.0)

Requirement already satisfied: tangled-up-in-unicode==0.1.0 in  
/srv/conda/envs/notebook/lib/python3.7/site-packages (from  
pandas\_profiling[notebook]) (0.1.0)

Requirement already satisfied: PyYAML>=5.0.0 in  
/srv/conda/envs/notebook/lib/python3.7/site-packages (from  
pandas\_profiling[notebook]) (6.0)

Requirement already satisfied: pydantic>=1.8.1 in  
/srv/conda/envs/notebook/lib/python3.7/site-packages (from  
pandas\_profiling[notebook]) (1.9.0)

Requirement already satisfied: jupyter-client>=6.0.0 in  
/srv/conda/envs/notebook/lib/python3.7/site-packages (from  
pandas\_profiling[notebook]) (7.1.0)

Requirement already satisfied: jupyter-core>=4.6.3 in  
/srv/conda/envs/notebook/lib/python3.7/site-packages (from  
pandas\_profiling[notebook]) (4.9.1)

Requirement already satisfied: networkx>=2.4 in  
/srv/conda/envs/notebook/lib/python3.7/site-packages (from  
visions[type\_image\_path]==0.7.4->pandas\_profiling[notebook]) (2.6.3)

Requirement already satisfied: attrs>=19.3.0 in  
/srv/conda/envs/notebook/lib/python3.7/site-packages (from  
visions[type\_image\_path]==0.7.4->pandas\_profiling[notebook]) (21.2.0)

Requirement already satisfied: imagehash in  
/srv/conda/envs/notebook/lib/python3.7/site-packages (from  
visions[type\_image\_path]==0.7.4->pandas\_profiling[notebook]) (4.2.1)

Requirement already satisfied: Pillow in  
/srv/conda/envs/notebook/lib/python3.7/site-packages (from  
visions[type\_image\_path]==0.7.4->pandas\_profiling[notebook]) (9.0.0)

Requirement already satisfied: scikit-learn in  
/srv/conda/envs/notebook/lib/python3.7/site-packages (from sklearn) (1.0.2)

Requirement already satisfied: six in  
/srv/conda/envs/notebook/lib/python3.7/site-packages (from catboost) (1.16.0)

Requirement already satisfied: graphviz in  
/srv/conda/envs/notebook/lib/python3.7/site-packages (from catboost) (0.19.1)

Requirement already satisfied: plotly in  
/srv/conda/envs/notebook/lib/python3.7/site-packages (from catboost) (5.5.0)

Requirement already satisfied: cycler>=0.10 in  
/srv/conda/envs/notebook/lib/python3.7/site-packages (from matplotlib) (0.11.0)

Requirement already satisfied: packaging>=20.0 in  
/srv/conda/envs/notebook/lib/python3.7/site-packages (from matplotlib) (21.3)

Requirement already satisfied: kiwisolver>=1.0.1 in  
/srv/conda/envs/notebook/lib/python3.7/site-packages (from matplotlib) (1.3.2)

Requirement already satisfied: pyparsing>=2.2.1 in  
/srv/conda/envs/notebook/lib/python3.7/site-packages (from matplotlib) (3.0.6)

Requirement already satisfied: fonttools>=4.22.0 in  
/srv/conda/envs/notebook/lib/python3.7/site-packages (from matplotlib) (4.28.5)

Requirement already satisfied: ipykernel>=4.5.1 in  
/srv/conda/envs/notebook/lib/python3.7/site-packages (from ipywidgets) (6.6.0)

Requirement already satisfied: jupyterlab-widgets>=1.0.0 in  
/srv/conda/envs/notebook/lib/python3.7/site-packages (from ipywidgets) (1.0.2)

Requirement already satisfied: ipython>=4.0.0 in  
/srv/conda/envs/notebook/lib/python3.7/site-packages (from ipywidgets) (7.30.1)

Requirement already satisfied: ipython-genutils~=0.2.0 in  
/srv/conda/envs/notebook/lib/python3.7/site-packages (from ipywidgets) (0.2.0)

Requirement already satisfied: widgetsnbextension~=3.5.0 in  
/srv/conda/envs/notebook/lib/python3.7/site-packages (from ipywidgets) (3.5.2)

Requirement already satisfied: nbformat>=4.2.0 in  
/srv/conda/envs/notebook/lib/python3.7/site-packages (from ipywidgets) (5.1.3)

Requirement already satisfied: traitlets>=4.3.1 in  
/srv/conda/envs/notebook/lib/python3.7/site-packages (from ipywidgets) (5.1.1)

Requirement already satisfied: importlib-metadata<5 in  
/srv/conda/envs/notebook/lib/python3.7/site-packages (from  
ipykernel>=4.5.1->ipywidgets) (4.10.0)

Requirement already satisfied: matplotlib-inline<0.2.0,>=0.1.0 in  
/srv/conda/envs/notebook/lib/python3.7/site-packages (from  
ipykernel>=4.5.1->ipywidgets) (0.1.3)

Requirement already satisfied: argcomplete>=1.12.3 in  
/srv/conda/envs/notebook/lib/python3.7/site-packages (from  
ipykernel>=4.5.1->ipywidgets) (1.12.3)

Requirement already satisfied: tornado<7.0,>=4.2 in  
/srv/conda/envs/notebook/lib/python3.7/site-packages (from  
ipykernel>=4.5.1->ipywidgets) (6.1)

Requirement already satisfied: debugpy<2.0,>=1.0.0 in

```

/srv/conda/envs/notebook/lib/python3.7/site-packages (from
ipykernel>=4.5.1->ipywidgets) (1.5.1)
Requirement already satisfied: backcall in
/srv/conda/envs/notebook/lib/python3.7/site-packages (from
ipython>=4.0.0->ipywidgets) (0.2.0)
Requirement already satisfied: pygments in
/srv/conda/envs/notebook/lib/python3.7/site-packages (from
ipython>=4.0.0->ipywidgets) (2.10.0)
Requirement already satisfied: pexpect>4.3 in
/srv/conda/envs/notebook/lib/python3.7/site-packages (from
ipython>=4.0.0->ipywidgets) (4.8.0)
Requirement already satisfied: setuptools>=18.5 in
/srv/conda/envs/notebook/lib/python3.7/site-packages (from
ipython>=4.0.0->ipywidgets) (60.0.4)
Requirement already satisfied: prompt-toolkit!=3.0.0,!3.0.1,<3.1.0,>=2.0.0 in
/srv/conda/envs/notebook/lib/python3.7/site-packages (from
ipython>=4.0.0->ipywidgets) (3.0.24)
Requirement already satisfied: pickleshare in
/srv/conda/envs/notebook/lib/python3.7/site-packages (from
ipython>=4.0.0->ipywidgets) (0.7.5)
Requirement already satisfied: decorator in
/srv/conda/envs/notebook/lib/python3.7/site-packages (from
ipython>=4.0.0->ipywidgets) (5.1.0)
Requirement already satisfied: jedi>=0.16 in
/srv/conda/envs/notebook/lib/python3.7/site-packages (from
ipython>=4.0.0->ipywidgets) (0.18.1)
Requirement already satisfied: nest-asyncio>=1.5 in
/srv/conda/envs/notebook/lib/python3.7/site-packages (from jupyter-
client>=6.0.0->pandas_profiling[notebook]) (1.5.4)
Requirement already satisfied: pyzmq>=13 in
/srv/conda/envs/notebook/lib/python3.7/site-packages (from jupyter-
client>=6.0.0->pandas_profiling[notebook]) (22.3.0)
Requirement already satisfied: entrypoints in
/srv/conda/envs/notebook/lib/python3.7/site-packages (from jupyter-
client>=6.0.0->pandas_profiling[notebook]) (0.3)
Requirement already satisfied: jsonschema!=2.5.0,>=2.4 in
/srv/conda/envs/notebook/lib/python3.7/site-packages (from
nbformat>=4.2.0->ipywidgets) (4.3.2)
Requirement already satisfied: typing-extensions>=3.7.4.3 in
/srv/conda/envs/notebook/lib/python3.7/site-packages (from
pydantic>=1.8.1->pandas_profiling[notebook]) (4.0.1)
Requirement already satisfied: idna<4,>=2.5 in
/srv/conda/envs/notebook/lib/python3.7/site-packages (from
requests>=2.24.0->pandas_profiling[notebook]) (3.1)
Requirement already satisfied: certifi>=2017.4.17 in
/srv/conda/envs/notebook/lib/python3.7/site-packages (from
requests>=2.24.0->pandas_profiling[notebook]) (2021.10.8)
Requirement already satisfied: charset-normalizer~2.0.0 in

```

```

/srv/conda/envs/notebook/lib/python3.7/site-packages (from
requests>=2.24.0->pandas_profiling[notebook]) (2.0.9)
Requirement already satisfied: urllib3<1.27,>=1.21.1 in
/srv/conda/envs/notebook/lib/python3.7/site-packages (from
requests>=2.24.0->pandas_profiling[notebook]) (1.26.7)
Requirement already satisfied: notebook>=4.4.1 in
/srv/conda/envs/notebook/lib/python3.7/site-packages (from
widgetsnbextension~=3.5.0->ipywidgets) (6.3.0)
Requirement already satisfied: tenacity>=6.2.0 in
/srv/conda/envs/notebook/lib/python3.7/site-packages (from plotly->catboost)
(8.0.1)
Requirement already satisfied: threadpoolctl>=2.0.0 in
/srv/conda/envs/notebook/lib/python3.7/site-packages (from scikit-
learn->sklearn) (3.0.0)
Requirement already satisfied: zipp>=0.5 in
/srv/conda/envs/notebook/lib/python3.7/site-packages (from importlib-
metadata<5->ipykernel>=4.5.1->ipywidgets) (3.6.0)
Requirement already satisfied: parso<0.9.0,>=0.8.0 in
/srv/conda/envs/notebook/lib/python3.7/site-packages (from
jedi>=0.16->ipython>=4.0.0->ipywidgets) (0.8.3)
Requirement already satisfied: pyrsistent!=0.17.0,!0.17.1,!0.17.2,>=0.14.0 in
/srv/conda/envs/notebook/lib/python3.7/site-packages (from
jsonschema!=2.5.0,>=2.4->nbformat>=4.2.0->ipywidgets) (0.18.0)
Requirement already satisfied: importlib-resources>=1.4.0 in
/srv/conda/envs/notebook/lib/python3.7/site-packages (from
jsonschema!=2.5.0,>=2.4->nbformat>=4.2.0->ipywidgets) (5.4.0)
Requirement already satisfied: Send2Trash>=1.5.0 in
/srv/conda/envs/notebook/lib/python3.7/site-packages (from
notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets) (1.8.0)
Requirement already satisfied: prometheus-client in
/srv/conda/envs/notebook/lib/python3.7/site-packages (from
notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets) (0.12.0)
Requirement already satisfied: argon2-cffi in
/srv/conda/envs/notebook/lib/python3.7/site-packages (from
notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets) (21.1.0)
Requirement already satisfied: terminado>=0.8.3 in
/srv/conda/envs/notebook/lib/python3.7/site-packages (from
notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets) (0.12.1)
Requirement already satisfied: nbconvert in
/srv/conda/envs/notebook/lib/python3.7/site-packages (from
notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets) (6.0.7)
Requirement already satisfied: ptyprocess>=0.5 in
/srv/conda/envs/notebook/lib/python3.7/site-packages (from
pexpect>4.3->ipython>=4.0.0->ipywidgets) (0.7.0)
Requirement already satisfied: wcwidth in
/srv/conda/envs/notebook/lib/python3.7/site-packages (from prompt-
toolkit!=3.0.0,!3.0.1,<3.1.0,>=2.0.0->ipython>=4.0.0->ipywidgets) (0.2.5)
Requirement already satisfied: PyWavelets in

```

```

/srv/conda/envs/notebook/lib/python3.7/site-packages (from
imagehash->visions[type_image_path]==0.7.4->pandas_profiling[notebook]) (1.2.0)
Requirement already satisfied: cffi>=1.0.0 in
/srv/conda/envs/notebook/lib/python3.7/site-packages (from
argon2-cffi->notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets) (1.15.0)
Requirement already satisfied: bleach in
/srv/conda/envs/notebook/lib/python3.7/site-packages (from
nbconvert->notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets) (4.1.0)
Requirement already satisfied: mistune<2,>=0.8.1 in
/srv/conda/envs/notebook/lib/python3.7/site-packages (from
nbconvert->notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets) (0.8.4)
Requirement already satisfied: nbclient<0.6.0,>=0.5.0 in
/srv/conda/envs/notebook/lib/python3.7/site-packages (from
nbconvert->notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets) (0.5.9)
Requirement already satisfied: testpath in
/srv/conda/envs/notebook/lib/python3.7/site-packages (from
nbconvert->notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets) (0.5.0)
Requirement already satisfied: pandocfilters>=1.4.1 in
/srv/conda/envs/notebook/lib/python3.7/site-packages (from
nbconvert->notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets) (1.5.0)
Requirement already satisfied: defusedxml in
/srv/conda/envs/notebook/lib/python3.7/site-packages (from
nbconvert->notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets) (0.7.1)
Requirement already satisfied: jupyterlab-pygments in
/srv/conda/envs/notebook/lib/python3.7/site-packages (from
nbconvert->notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets) (0.1.2)
Requirement already satisfied: pycparser in
/srv/conda/envs/notebook/lib/python3.7/site-packages (from cffi>=1.0.0->argon2-
cffi->notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets) (2.21)
Requirement already satisfied: webencodings in
/srv/conda/envs/notebook/lib/python3.7/site-packages (from
bleach->nbconvert->notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets)
(0.5.1)

```

## 1.2 Dealing With Google Colab Issues

```
[2]: #from google.colab import output
      #output.enable_custom_widget_manager()
```

```
[3]: !jupyter nbextension enable --py widgetsnbextension
```

```

Enabling notebook extension jupyter-js-widgets/extension...
  - Validating: OK

```

```
[4]: import matplotlib
      import matplotlib.pyplot as plt
```

```
[5]: %matplotlib inline
```

### 1.3 Data Acquisition

```
[6]: !wget https://archive.ics.uci.edu/ml/machine-learning-databases/mushroom/  
      ↪agaricus-lepiota.data
```

```
--2022-01-19 08:24:58-- https://archive.ics.uci.edu/ml/machine-learning-  
databases/mushroom/agaricus-lepiota.data  
Resolving archive.ics.uci.edu (archive.ics.uci.edu)... 128.195.10.252  
Connecting to archive.ics.uci.edu (archive.ics.uci.edu)|128.195.10.252|:443...  
connected.  
HTTP request sent, awaiting response... 200 OK  
Length: 373704 (365K) [application/x-httpd-php]  
Saving to: 'agaricus-lepiota.data.3'
```

```
agaricus-lepiota.da 100%[=====>] 364.95K  1.39MB/s    in 0.3s
```

```
2022-01-19 08:24:58 (1.39 MB/s) - 'agaricus-lepiota.data.3' saved  
[373704/373704]
```

```
[7]: !wget https://archive.ics.uci.edu/ml/machine-learning-databases/mushroom/  
      ↪agaricus-lepiota.names
```

```
--2022-01-19 08:24:59-- https://archive.ics.uci.edu/ml/machine-learning-  
databases/mushroom/agaricus-lepiota.names  
Resolving archive.ics.uci.edu (archive.ics.uci.edu)... 128.195.10.252  
Connecting to archive.ics.uci.edu (archive.ics.uci.edu)|128.195.10.252|:443...  
connected.  
HTTP request sent, awaiting response... 200 OK  
Length: 6816 (6.7K) [application/x-httpd-php]  
Saving to: 'agaricus-lepiota.names.3'
```

```
agaricus-lepiota.na 100%[=====>]    6.66K  --.-KB/s    in 0s
```

```
2022-01-19 08:24:59 (154 MB/s) - 'agaricus-lepiota.names.3' saved [6816/6816]
```

```
[8]: # Dataset Description
```

```
!cat agaricus-lepiota.names
```

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 dissertation, 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), Bruges, Belgium 16-18.4.1997, pp. xx-xx

Włodzisław Duch, Department of Computer Methods, Nicholas Copernicus University, 87-100 Toruń, Grudziadzka 5, Poland  
e-mail: [duch@phys.uni.torun.pl](mailto:duch@phys.uni.torun.pl)  
WWW <http://www.phys.uni.torun.pl/kmk/>

Date: Mon, 17 Feb 1997 13:47:40 +0100  
From: Włodzisław Duch <[duch@phys.uni.torun.pl](mailto: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

```

[9]: *# Dataset Pre-view*

```
!head agaricus-lepiota.data
```

```
p,x,s,n,t,p,f,c,n,k,e,e,s,s,w,w,p,w,o,p,k,s,u  
e,x,s,y,t,a,f,c,b,k,e,c,s,s,w,w,p,w,o,p,n,n,g  
e,b,s,w,t,l,f,c,b,n,e,c,s,s,w,w,p,w,o,p,n,n,m  
p,x,y,w,t,p,f,c,n,n,e,e,s,s,w,w,p,w,o,p,k,s,u  
e,x,s,g,f,n,f,w,b,k,t,e,s,s,w,w,p,w,o,e,n,a,g  
e,x,y,y,t,a,f,c,b,n,e,c,s,s,w,w,p,w,o,p,k,n,g  
e,b,s,w,t,a,f,c,b,g,e,c,s,s,w,w,p,w,o,p,k,n,m  
e,b,y,w,t,l,f,c,b,n,e,c,s,s,w,w,p,w,o,p,n,s,m  
p,x,y,w,t,p,f,c,n,p,e,e,s,s,w,w,p,w,o,p,k,v,g  
e,b,s,y,t,a,f,c,b,g,e,c,s,s,w,w,p,w,o,p,k,s,m
```

## 1.4 Data Loading

```
[10]: import numpy as np  
import pandas as pd  
from pandas_profiling import ProfileReport
```

```
[11]: df = pd.read_csv('agaricus-lepiota.data', header=None)
```

```
[12]: colnames = ["class",  
"cap-shape",  
"cap-surface",  
"cap-color",  
"bruises?",  
"odor",  
"gill-attachment",  
"gill-spacing",  
"gill-size",  
"gill-color",  
"stalk-shape",  
"stalk-root",  
"stalk-surface-above-ring",  
"stalk-surface-below-ring",  
"stalk-color-above-ring",  
"stalk-color-below-ring",  
"veil-type",  
"veil-color",  
"ring-number",  
"ring-type",  
"spore-print-color",  
"population",  
"habitat"  
]
```

```
[13]: df.columns = colnames
```

```
[14]: df['stalk-root'].replace('?', np.nan, inplace=True)
```

```
[15]: df.head()
```

```
[15]: class cap-shape cap-surface cap-color bruises? odor gill-attachment \
0      p          x          s          n          t          p          f
1      e          x          s          y          t          a          f
2      e          b          s          w          t          l          f
3      p          x          y          w          t          p          f
4      e          x          s          g          f          n          f

    gill-spacing gill-size gill-color ... stalk-surface-below-ring \
0              c          n          k ...                          s
1              c          b          k ...                          s
2              c          b          n ...                          s
3              c          n          n ...                          s
4              w          b          k ...                          s

    stalk-color-above-ring stalk-color-below-ring veil-type veil-color \
0                          w                          w          p          w
1                          w                          w          p          w
2                          w                          w          p          w
3                          w                          w          p          w
4                          w                          w          p          w

    ring-number ring-type spore-print-color population habitat
0              o          p                  k          s          u
1              o          p                  n          n          g
2              o          p                  n          n          m
3              o          p                  k          s          u
4              o          e                  n          a          g

[5 rows x 23 columns]
```

## 1.5 Exploratory Data Analysis

**Note:** Pre-processing step is skipped as all values are categorical and catboost can deal with missing values on its own.

```
[16]: profile = ProfileReport(df, title="Exploratory Data Analysis Report",
    ↪explorative=True)
```

```
[ ]: profile.to_notebook_iframe()
```

Summarize dataset: 0% | 0/5 [00:00<?, ?it/s]

## 1.6 Data Preparation

### 1.6.1 Data Formatting

```
[ ]: # Converting NaN values to string so that catboost can deal with it

df['stalk-root'].replace(np.nan, 'Nan', inplace=True)
```

### 1.6.2 Data Splitting

```
[ ]: from sklearn.model_selection import train_test_split

[ ]: df_train_val, df_test = train_test_split(df, test_size=0.2, random_state=42)

[ ]: df_train, df_val = train_test_split(df_train_val, test_size=0.125,
    ↪ random_state=42)

[ ]: X_train, y_train = df_train.drop('class', axis=1), df_train['class']

[ ]: X_test, y_test = df_test.drop('class', axis=1), df_test['class']

[ ]: X_val, y_val = df_val.drop('class', axis=1), df_val['class']
```

### 1.6.3 Data Pool Creation

```
[ ]: import catboost as cb

[ ]: categorical_indicies = colnames[1:]

[ ]: train_dataset = cb.Pool(X_train, y_train,
    cat_features=categorical_indicies)

[ ]: val_dataset = cb.Pool(X_val, y_val,
    cat_features=categorical_indicies)

[ ]: test_dataset = cb.Pool(X_test, y_test,
    cat_features=categorical_indicies)
```

## 1.7 Training

```
[ ]: model = cb.CatBoostClassifier(depth=8,
    iterations=100,
    eval_metric='Accuracy',
    boosting_type="Ordered",
    bagging_temperature=0,
    use_best_model=True,
    loss_function='Logloss')

[ ]: model.fit(train_dataset, eval_set=val_dataset, plot=True)
```

## 1.8 Evaluation

```
[ ]: # Compute metrics for all models

model.eval_metrics(test_dataset, ['Accuracy', 'Logloss', 'AUC', 'CrossEntropy',
    ↪ 'Recall', 'Precision', 'F1', 'BalancedAccuracy'])

[ ]: import seaborn as sns

[ ]: def plot_feature_importance(importance, names, model_type):
    feature_importance = np.array(importance)
    feature_names = np.array(names)
    data={'feature_names':feature_names,
          'feature_importance':feature_importance}
    fi_df = pd.DataFrame(data)
    fi_df.sort_values(by=['feature_importance'],
                      ascending=False,inplace=True)
    plt.figure(figsize=(10,8))
    sns.barplot(x=fi_df['feature_importance'],
                y=fi_df['feature_names'])
    plt.title(model_type + ' FEATURE IMPORTANCE')
    plt.xlabel('FEATURE IMPORTANCE')
    plt.ylabel('FEATURE NAMES')
    plt.show()

[ ]: %matplotlib inline

[ ]: plot_feature_importance(model.get_feature_importance(), categorical_indicies,
    ↪ 'CATBOOST')

[ ]: pred = model.predict(X_test)

[ ]: pred_proba = model.predict_proba(X_test)

[ ]: from sklearn.metrics import classification_report, confusion_matrix, roc_curve,
    ↪ RocCurveDisplay

[ ]: print(classification_report(y_test, pred))

[ ]: sns.heatmap(confusion_matrix(y_test, pred), annot=True)
plt.show()

[ ]: RocCurveDisplay.from_predictions(y_test, pred_proba[:, 0], pos_label='e')
plt.show()

[ ]:
```