MS Azure ML Designer를 활용한 개인 실습

Lab : 버섯 데이터

강명호

Mushroom 데이터

Mushroom
Donated on 4/26/1987

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

Dataset Characteristics
Multivariate

Subject Area
Biology

Associated Tasks
Classification

Feature Type
Categorical

 # Instances
8124

 # Features
22

Dataset Information

Additional 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...

SHOW MORE ⌄

Has Missing Values?
Yes

Mushroom 데이터

- 23종의 주름버섯 데이터
- 버섯의 식용 구분 : 식용 가능한 버섯, 독이 있는 버섯, 또는 식용 여부가 불확실한 버섯
  - 데이터세트에서는 식용 여부가 불확실한 버섯은 독이 있는 버섯의 범주에 포함시킴

- → 독이 있는지 (p : poisonous), 없는지(e : edible)로 버섯을 구분하는 분류 문제

Dataset Information

Additional 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.

SHOW LESS ∧

Has Missing Values?
Yes
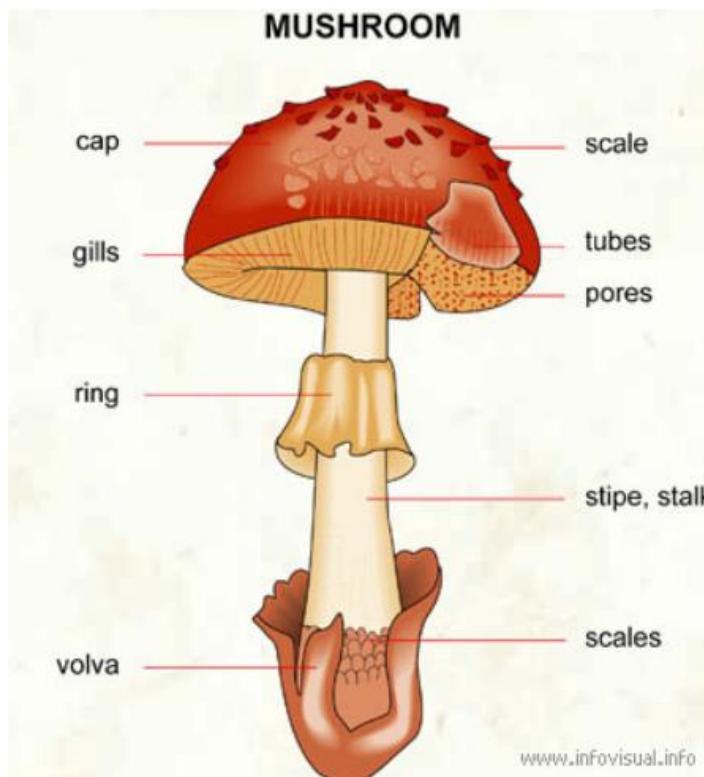
Mushroom 데이터

Variables Table

Variable Name	Role	Type	Description	Units	Missing Values
poisonous	Target	Categorical			no
cap-shape	Feature	Categorical	bell=b,conical=c,convex=x,flat=f, knobbed=k,sunken=s		no
cap-surface	Feature	Categorical	fibrous=f,grooves=g,scaly=y,smooth=s		no
cap-color	Feature	Binary	brown=n,buff=b,cinnamon=c,gray=g,green=r, pink=p,purple=u,red=e,white=w,yellow=y		no
bruises	Feature	Categorical	bruises=t,no=f		no
odor	Feature	Categorical	almond=a,anise=l,creosote=c,fishy=y,foul=f, musty=m,none=n,pungent=p,spicy=s		no
gill-attachment	Feature	Categorical	attached=a,descending=d,free=f,notched=n		no
gill-spacing	Feature	Categorical	close=c,crowded=w,distant=d		no
gill-size	Feature	Categorical	broad=b,narrow=n		no
gill-color	Feature	Categorical	black=k,brown=n,buff=b,chocolate=h,gray=g, green=r,orange=o,pink=p,purple=u,red=e, white=w,yellow=y		no

Mushroom 데이터

MUSHROOM

Variables Table

Variable Name	Role	Type	Description	Units	Missing Value
stalk-shape	Feature	Categorical	enlarging=e,tapering=t		no
stalk-root	Feature	Categorical	bulbous=b,club=c,cup=u,equal=e, rhizomorphs=z,rooted=r,missing=?		yes
stalk-surface-above-ring	Feature	Categorical	fibrous=f,scaly=y,silky=k,smooth=s		no
stalk-surface-below-ring	Feature	Categorical	fibrous=f,scaly=y,silky=k,smooth=s		no
stalk-color-above-ring	Feature	Categorical	brown=n,buff=b,cinnamon=c,gray=g,orange=o, pink=p,red=e,white=w,yellow=y		no
stalk-color-below-ring	Feature	Categorical	brown=n,buff=b,cinnamon=c,gray=g,orange=o, pink=p,red=e,white=w,yellow=y		no
veil-type	Feature	Binary	partial=p,universal=u		no
veil-color	Feature	Categorical	brown=n,orange=o,white=w,yellow=y		no
ring-number	Feature	Categorical	none=n,one=o,two=t		no
ring-type	Feature	Categorical	cobwebby=c,evanescent=e,flaring=f,large=l, none=n,pendant=p,sheathing=s,zone=z		no

Mushroom 데이터

Variables Table

Variable Name	Role	Type	Description	Units	Missing Values
spore-print-color	Feature	Categorical	black=k,brown=n,buff=b,chocolate=h,green=r, orange=o,purple=u,white=w,yellow=y		no
population	Feature	Categorical	abundant=a,clustered=c,numerous=n, scattered=s,several=v,solitary=y		no
habitat	Feature	Categorical	grasses=g,leaves=l,meadows=m,paths=p, urban=u,waste=w,woods=d		no

Mushroom 데이터

agaricus-lepiota.data file

```
1   p,x,s,n,t,p,f,c,n,k,e,e,s,s,w,w,p,w,o,p,k,s,u
2   e,x,s,y,t,a,f,c,b,k,e,c,s,s,w,w,p,w,o,p,n,n,g
3   e,b,s,w,t,l,f,c,b,n,e,c,s,s,w,w,p,w,o,p,n,n,m
4   p,x,y,w,t,p,f,c,n,n,e,e,s,s,w,w,p,w,o,p,k,s,u
5   e,x,s,g,f,n,f,w,b,k,t,e,s,s,w,w,p,w,o,e,n,a,g
6   e,x,y,y,t,a,f,c,b,n,e,c,s,s,w,w,p,w,o,p,k,n,g
7   e,b,s,w,t,a,f,c,b,g,e,c,s,s,w,w,p,w,o,p,k,n,m
8   e,b,y,w,t,l,f,c,b,n,e,c,s,s,w,w,p,w,o,p,n,s,m
9   p,x,y,w,t,p,f,c,n,p,e,e,s,s,w,w,p,w,o,p,k,v,g
10  e,b,s,y,t,a,f,c,b,g,e,c,s,s,w,w,p,w,o,p,k,s,m
11  e,x,y,y,t,l,f,c,b,g,e,c,s,s,w,w,p,w,o,p,n,n,g
12  e,x,y,y,t,a,f,c,b,n,e,c,s,s,w,w,p,w,o,p,k,s,m
13  e,b,s,y,t,a,f,c,b,w,e,c,s,s,w,w,p,w,o,p,n,s,g
14  p,x,y,w,t,p,f,c,n,k,e,e,s,s,w,w,p,w,o,p,n,v,u
15  e,x,f,n,f,n,f,w,b,n,t,e,s,f,w,w,p,w,o,e,k,a,g
16  e,s,f,g,f,n,f,c,n,k,e,e,s,s,w,w,p,w,o,p,n,y,u
17  e,f,f,w,f,n,f,w,b,k,t,e,s,s,w,w,p,w,o,e,n,a,g
18  p,x,s,n,t,p,f,c,n,n,e,e,s,s,w,w,p,w,o,p,k,s,g
19  p,x,y,w,t,p,f,c,n,n,e,e,s,s,w,w,p,w,o,p,n,s,u
20  p,x,s,n,t,p,f,c,n,k,e,e,s,s,w,w,p,w,o,p,n,s,u
```

Mushroom 데이터

agaricus-lepiota.names file

특성컬럼 이름 확인 가능

첫 컬럼 : classes

```
106  7. Attribute Information: (classes: edible=e, poisonous=p)
107        1. cap-shape:                bell=b,conical=c,convex=x,flat=f,
108                                      knobbed=k,sunken=s
109        2. cap-surface:              fibrous=f,grooves=g,scaly=y,smooth=s
110        3. cap-color:                brown=n,buff=b,cinnamon=c,gray=g,green=r,
111                                      pink=p,purple=u,red=e,white=w,yellow=y
112        4. bruises?:                 bruises=t,no=f
113        5. odor:                     almond=a,anise=l,creosote=c,fishy=y,foul=f,
114                                      musty=m,none=n,pungent=p,spicy=s
115        6. gill-attachment:          attached=a,descending=d,free=f,notched=n
116        7. gill-spacing:             close=c,crowded=w,distant=d
117        8. gill-size:                broad=b,narrow=n
118        9. gill-color:               black=k,brown=n,buff=b,chocolate=h,gray=g,
119                                      green=r,orange=o,pink=p,purple=u,red=e,
120                                      white=w,yellow=y
121       10. stalk-shape:              enlarging=e,tapering=t
122       11. stalk-root:               bulbous=b,club=c,cup=u,equal=e,
123                                      rhizomorphs=z,rooted=r,missing=?
124       12. stalk-surface-above-ring: fibrous=f,scaly=y,silky=k,smooth=s
125       13. stalk-surface-below-ring: fibrous=f,scaly=y,silky=k,smooth=s
126       14. stalk-color-above-ring:   brown=n,buff=b,cinnamon=c,gray=g,orange=o,
127                                      pink=p,red=e,white=w,yellow=y
128       15. stalk-color-below-ring:   brown=n,buff=b,cinnamon=c,gray=g,orange=o,
129                                      pink=p,red=e,white=w,yellow=y
130       16. veil-type:                partial=p,universal=u
131       17. veil-color:               brown=n,orange=o,white=w,yellow=y
132       18. ring-number:              none=n,one=o,two=t
133       19. ring-type:                cobwebby=c,evanescent=e,flaring=f,large=l,
134                                      none=n,pendant=p,sheathing=s,zone=z
135       20. spore-print-color:        black=k,brown=n,buff=b,chocolate=h,green=r,
136                                      orange=o,purple=u,white=w,yellow=y
137       21. population:               abundant=a,clustered=c,numerous=n,
138                                      scattered=s,several=v,solitary=y
139       22. habitat:                  grasses=g,leaves=l,meadows=m,paths=p,
140                                      urban=u,waste=w,woods=d
141
```

Mushroom 데이터

agaricus-lepiota.names file

- 누락값 : 2480개
  - "?" 로 표시
  - #11 컬럼에 있음
  - (Label 컬럼을
    포함하면 #12 컬럼)

- 범주 분포
  - edible : 51.8%
  - poisonous : 48.2%

```
142   8. Missing Attribute Values: 2480 of them (denoted by "?"), all for
143      attribute #11.
144
145   9. Class Distribution:
146      --      edible: 4208 (51.8%)
147      -- poisonous: 3916 (48.2%)
148      --       total: 8124 instances
149
```

추가 정보 : 하나의 범주값만 갖는 특성이 있음