DETAILED PROJECT REPORT

iNeuron

Mushroom Classifications

Safe to eat or deadly poison?

By Rohan Kadam And Vijay kamble

PROJECT DETAIL

Project Title	Mushroom Classification
Technologies	Machine Learning Technology
Domain	Agriculture
level	Intermediate

Objective

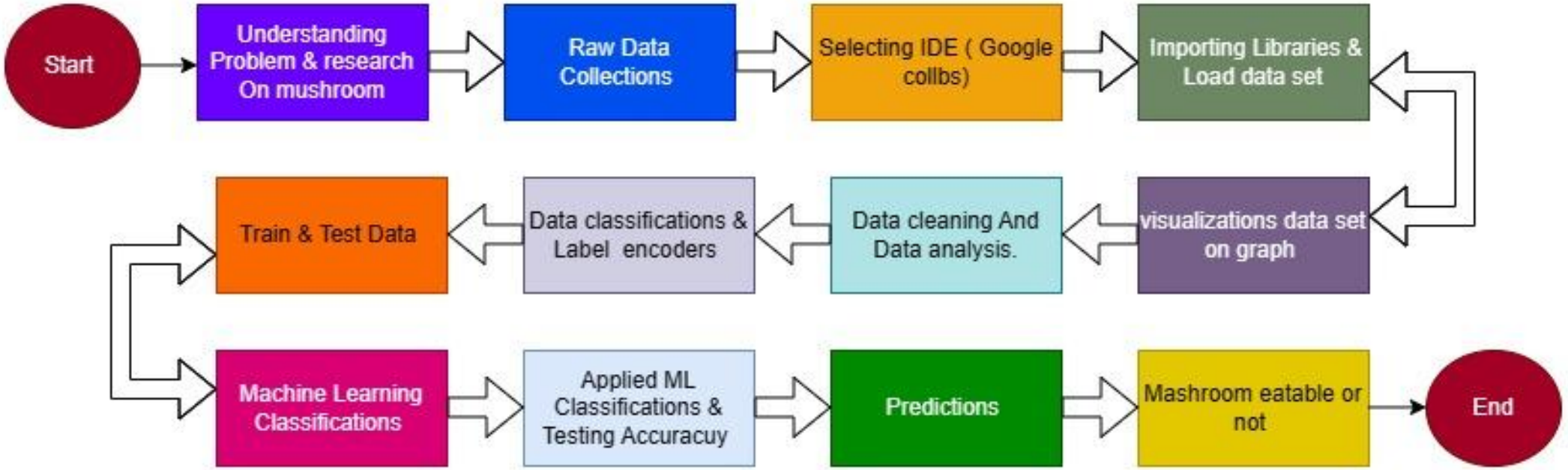
The main goal is to predict which mushroom is poisonous & which is edible

PROBLEM STATEMENT

The Audubon Society Field Guide to North American Mushrooms contains descriptions of hypothetical samples corresponding to 23 species of gilled mushrooms in the Agaricus and Lepiota Family Mushroom (1981). Each species is labelled as either definitely edible, definitely poisonous, or maybe edible but not recommended. This last category was merged with the toxic category. The Guide asserts unequivocally that there is no simple rule for judging a mushroom's edibility, such as "leaflets three, leave it be" for Poisonous Oak and Ivy.

The main goal is to predict which mushroom is poisonous & which is edible.

The Architecture



DATASET INFORMATION

The dataset used in this project contains 8124 instances of mushrooms with 23 features like cap-shape, cap-surface, cap-color, bruises, odor, etc.

Attribute Information:

(classes: edible=e, poisonous=p)

cap-shape: bell=b,conical=c,convex=x,flat=f, knobbed=k,sunken=s

cap-surface: fibrous=f,grooves=g,scaly=y,smooth=s

cap-color: brown=n,buff=b,cinnamon=c,gray=g,green=r,pink=p,purple=u,red=e,white=w,yellow=y

bruises: bruises=t,no=f

odor: almond=a,anise=l,creosote=c,fishy=y,foul=f,musty=m,none=n,pungent=p,spicy=s

gill-attachment: attached=a,descending=d,free=f,notched=n

gill-spacing: close=c,crowded=w,distant=d

gill-size: broad=b,narrow=n

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

stalk-shape: enlarging=e,tapering=t

stalk-root: bulbous=b,club=c,cup=u,equal=e,rhizomorphs=z,rooted=r,missing=?

stalk-surface-above-ring: fibrous=f,scaly=y,silky=k,smooth=s

stalk-surface-below-ring: fibrous=f,scaly=y,silky=k,smooth=s

stalk-color-above-ring: brown=n,buff=b,cinnamon=c,gray=g,orange=o,pink=p,red=e,white=w,yellow=y

stalk-color-below-ring: brown=n,buff=b,cinnamon=c,gray=g,orange=o,pink=p,red=e,white=w,yellow=y

veil-type: partial=p,universal=u

veil-color: brown=n,orange=o,white=w,yellow=y

ring-number : none=n,one=o,two=t

ring-type : cobwebby=c,evanescent=e,flaring=f,large=l,none=n,pendant=p,sheathing=s,zone=z

spore-print-color : black=k,brown=n,buff=b,chocolate=h,green=r,orange=o,purple=u,white=w,yellow=y

population: abundant=a,clustered=c,numerous=n,scattered=s,several=v,solitary=y

habitat: grasses=g,leaves=l,meadows=m,paths=p,urban=u,waste=w,woods=d

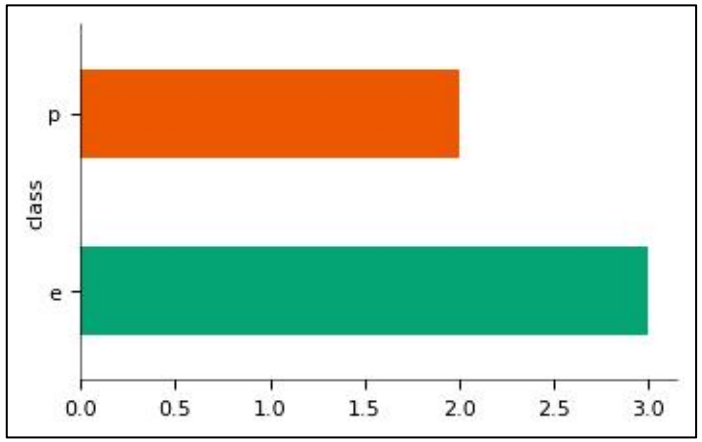
The data set is given in csv file Column and about data set as follows :

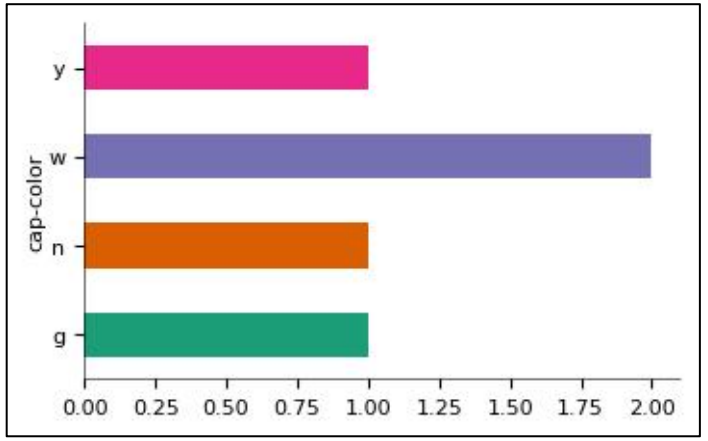
Data Set Characteristics:	Multivariate	**Number of Instances:**	8124	**Area:**	Life
Attribute Characteristics:	Categorical	**Number of Attributes:**	22	**Date Donated**	1987-04-27
Associated Tasks:	Classification	**Missing Values?**	Yes	**Number of Web Hits:**	37270

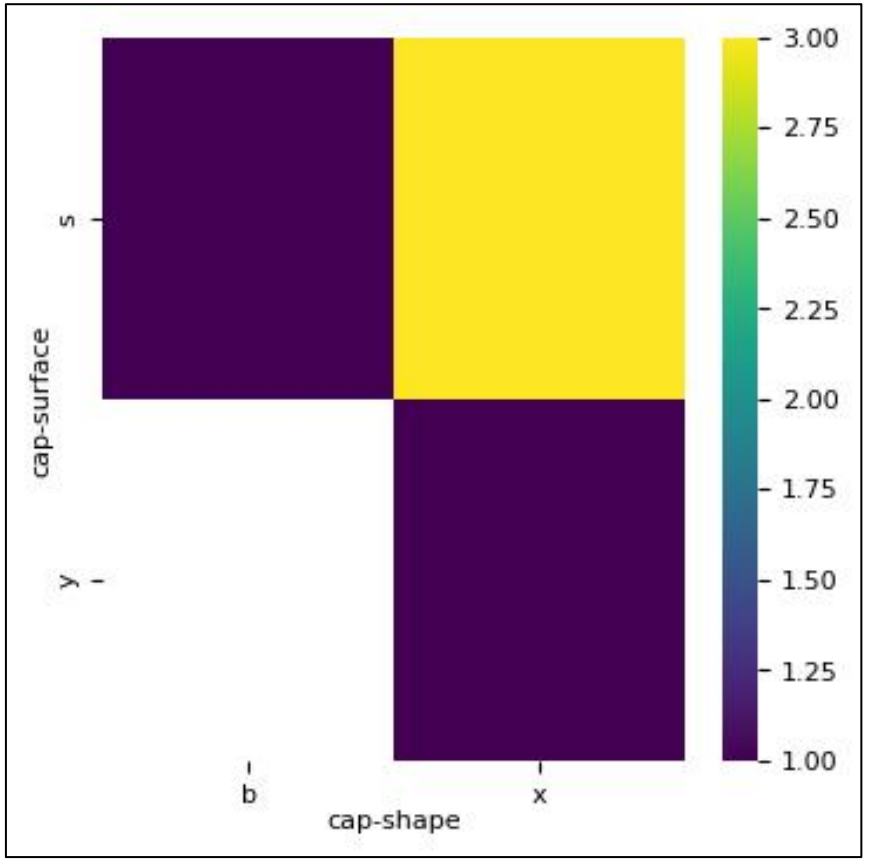
	class	cap-shape	cap-surface	cap-color	bruises	odor	gill-attachment	gill-spacing	gill-size	gill-color	...	stalk-surface-below-ring	stalk-color-above-ring	stalk-color-below-ring	veil-type	veil-color	ring-number	ring-type	spore-print-color	population	habitat
8119	e	k	s	n	f	n	a	c	b	y	...	s	o	o	p	o	o	p	b	c	l
8120	e	x	s	n	f	n	a	c	b	y	...	s	o	o	p	n	o	p	b	v	l
8121	e	f	s	n	f	n	a	c	b	n	...	s	o	o	p	o	o	p	b	c	l
8122	p	k	y	n	f	y	f	c	n	b	...	k	w	w	p	w	o	e	w	v	l
8123	e	x	s	n	f	n	a	c	b	y	...	s	o	o	p	o	o	p	o	c	l

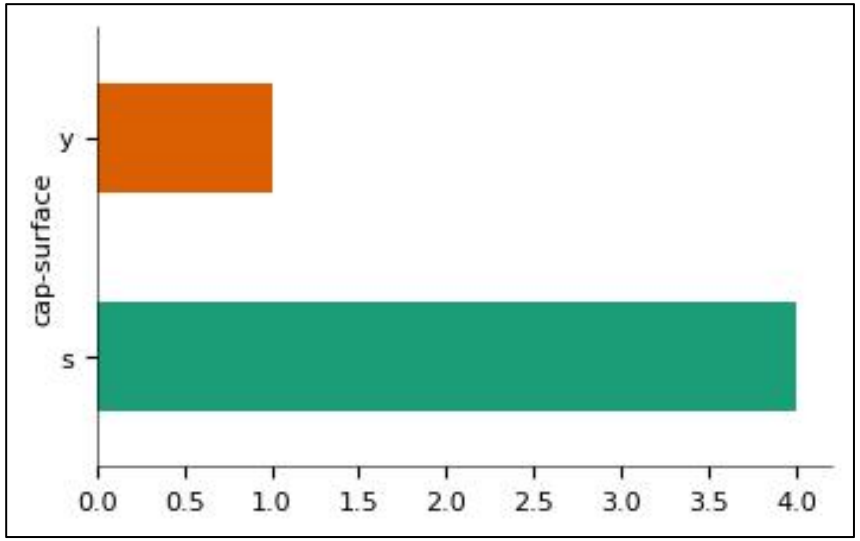
5 rows × 23 columns

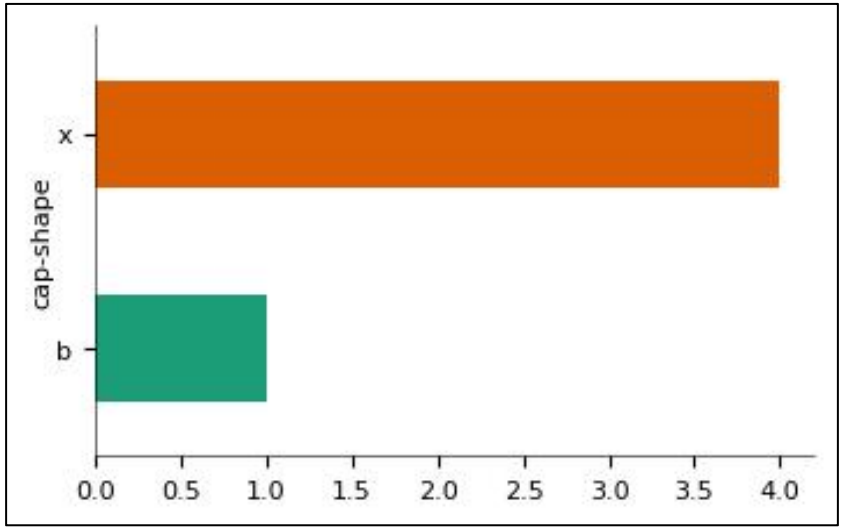
2-D Categorical Graphs

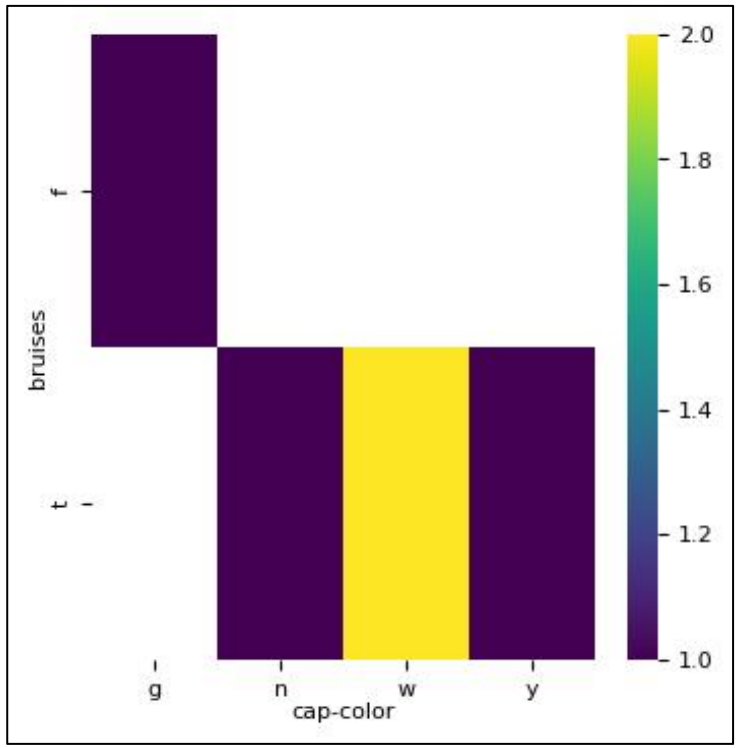


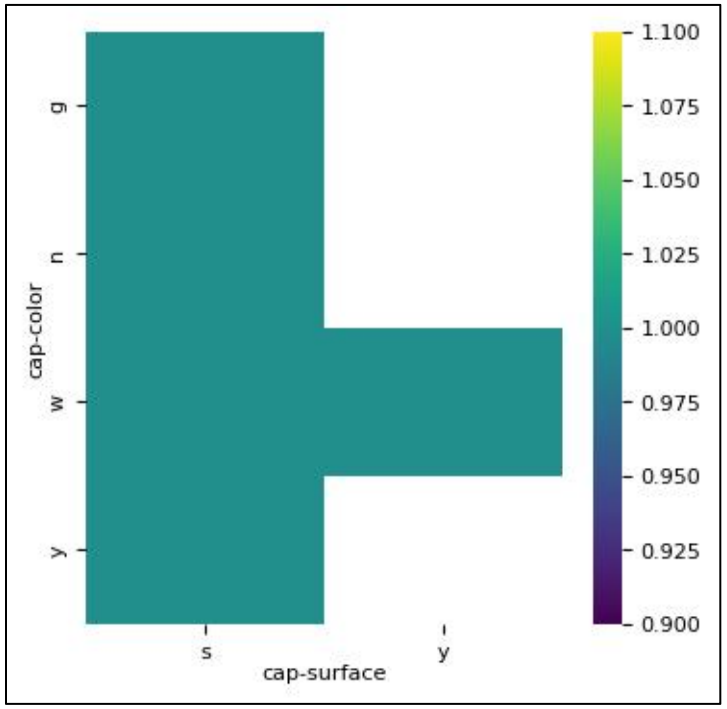


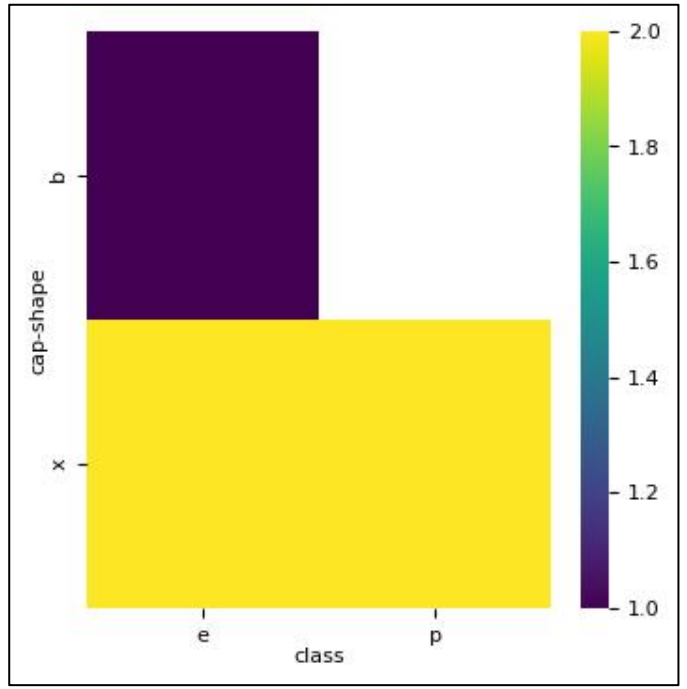


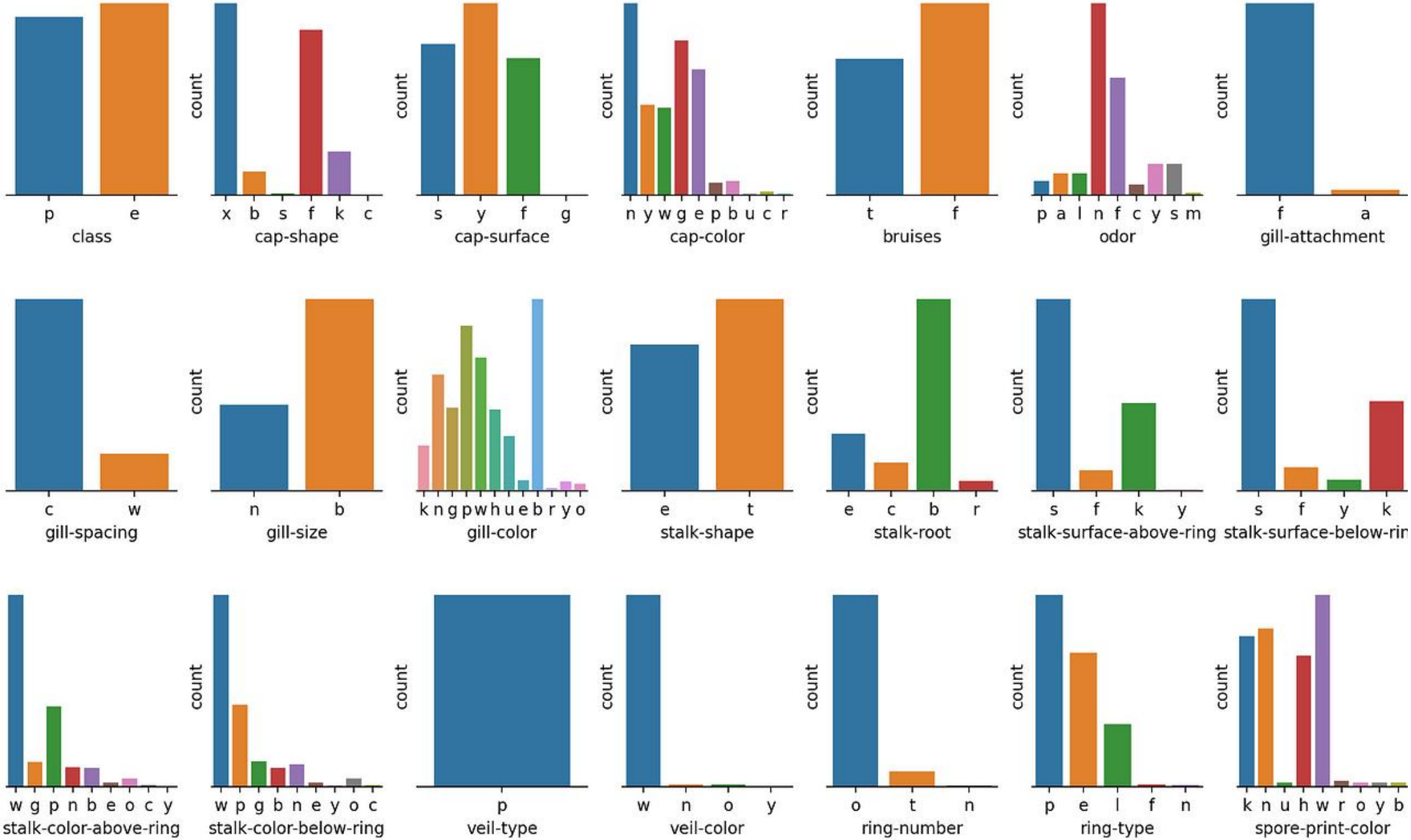



cap-shape
x
b
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0


bruises
f
t
g
n
w
y
cap-color
2.0
1.8
1.6
1.4
1.2
1.0


cap-color
g
n
w
y
s
y
cap-surface
1.100
1.075
1.050
1.025
1.000
0.975
0.950
0.925
0.900

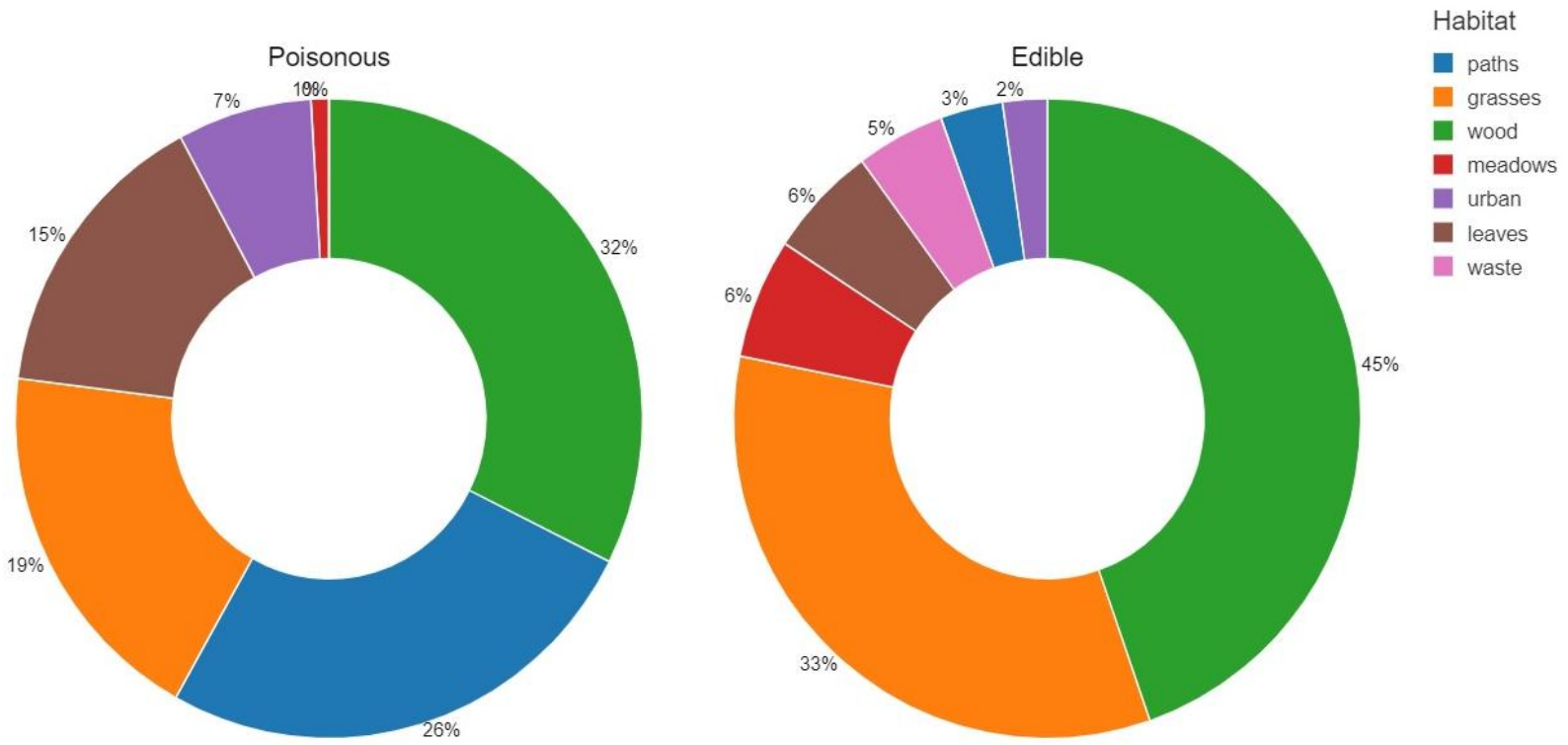

cap-shape
b
x
e
p
class
2.0
1.8
1.6
1.4
1.2
1.0


count
p e
class
x b s f k c
cap-shape
s y f g
cap-surface
n y w g e p b u c r
cap-color
t f
bruises
p a l n f c y s m
odor
f a
gill-attachment
c w
gill-spacing
n b
gill-size
k n g p w h u e b r y o
gill-color
e t
stalk-shape
e c b r
stalk-root
s f k y
stalk-surface-above-ring
s f y k
stalk-surface-below-rin
w g p n b e o c y
stalk-color-above-ring
w p g b n e y o c
stalk-color-below-ring
p
veil-type
w n o y
veil-color
o t n
ring-number
p e l f n
ring-type
k n u h w r o y b
spore-print-color

Poision And Edible Class

```
END AS habitat
from MushroomData group by habitat, class
```

▸ (2) Spark Jobs



Plot Options...

Command took 0.79 seconds -- by ybhavesh22@gmail.com at 2/21/2021, 1:24:51 PM on Spark Cluster

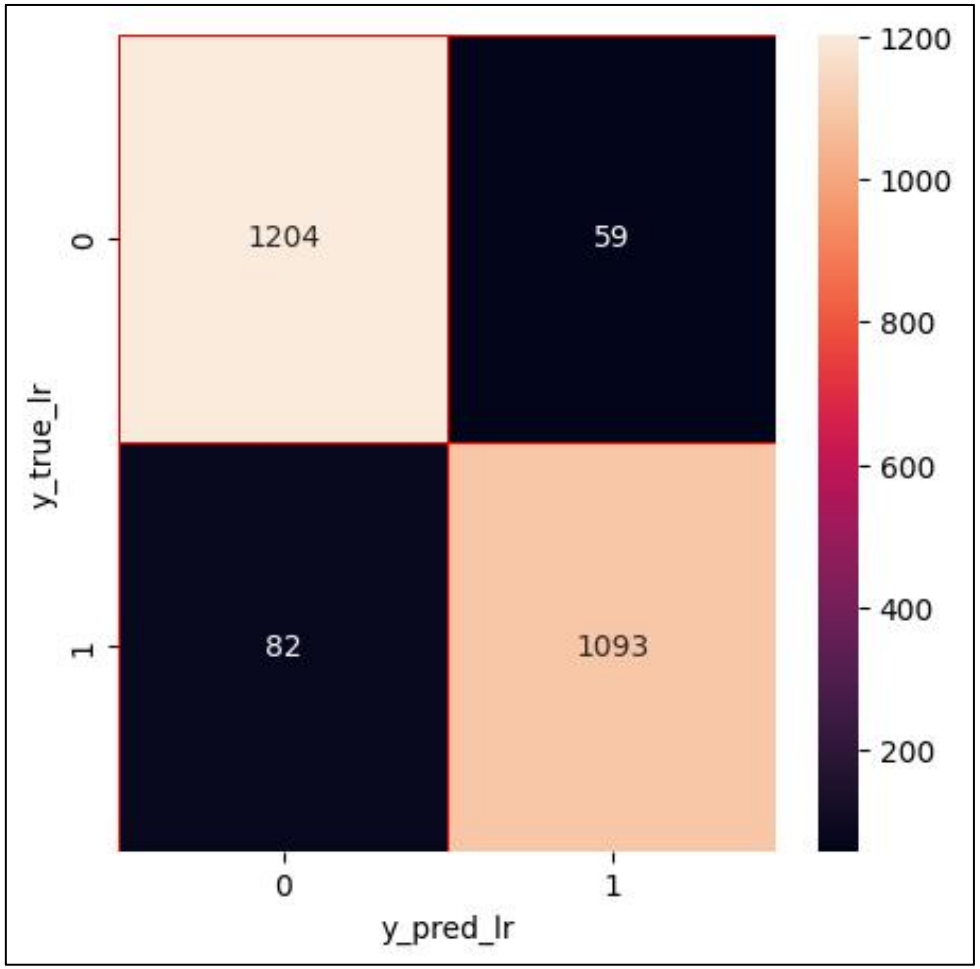
e here to search

ENG US 11:06 22-02-2021

Machine Learning Classification Test Result

Sr.no	Classifications	Result
1	#Random Forest Classification	100%
2	#Decision tree classification	100%
3	#naive bayes classfications	91.88%
4	#SVM Classification	100%
5	#K-Nearest neighbour	100%
6	#Logistic Regressions	94%

Confusion Matrix



Conclusion

From the confusion matrix, we saw that our train and test data is balanced.
Most of classfication methods hit 100% accuracy with this dataset.

In conclusion, the application of machine learning in mushroom classification has demonstrated its remarkable potential in automating and enhancing the accuracy of identifying mushroom species. Through the utilization of advanced algorithms and vast datasets, we have witnessed the development of robust models capable of distinguishing between edible and toxic mushrooms with a high degree of confidence. As technology continues to advance and more research is conducted, we can anticipate even greater strides in the accuracy and efficiency of mushroom classification using machine learning. This, in turn, will contribute to safer mushroom foraging practices, greater understanding of fungal biodiversity, and the preservation of ecosystems.

Thank You