

```
In [54]: from sklearn.ensemble import AdaBoostClassifier
from sklearn.preprocessing import LabelEncoder
from sklearn.tree import DecisionTreeClassifier
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn import metrics

df1 = pd.read_csv('C:/Users/USER/Desktop/MLESEMBLESDOCS-02NOV2021/mushroomdataset/mushrooms.csv')
```

```
In [55]: df1 = df1.sample(frac = 1)
```

```
In [56]: df1.columns
```

```
Out[56]: Index(['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'],
              dtype='object')
```

```
In [57]: for label in df1.columns:
          df1[label] = LabelEncoder().fit(df1[label]).transform(df1[label])
```

```
In [58]: df1.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 8124 entries, 7701 to 2463
Data columns (total 23 columns):
 #   Column                Non-Null Count  Dtype
---  -
 0   class                 8124 non-null  int32
 1   cap-shape             8124 non-null  int32
 2   cap-surface           8124 non-null  int32
 3   cap-color             8124 non-null  int32
 4   bruises               8124 non-null  int32
 5   odor                 8124 non-null  int32
 6   gill-attachment       8124 non-null  int32
 7   gill-spacing          8124 non-null  int32
 8   gill-size             8124 non-null  int32
 9   gill-color            8124 non-null  int32
10  stalk-shape           8124 non-null  int32
11  stalk-root            8124 non-null  int32
```

```

12 stalk-surface-above-ring 8124 non-null int32
13 stalk-surface-below-ring 8124 non-null int32
14 stalk-color-above-ring 8124 non-null int32
15 stalk-color-below-ring 8124 non-null int32
16 veil-type 8124 non-null int32
17 veil-color 8124 non-null int32
18 ring-number 8124 non-null int32
19 ring-type 8124 non-null int32
20 spore-print-color 8124 non-null int32
21 population 8124 non-null int32
22 habitat 8124 non-null int32
dtypes: int32(23)
memory usage: 793.4 KB

```

In [59]: `df1.shape`

Out[59]: (8124, 23)

```

In [60]: X = df1.drop(['class'], axis = 1)
        Y = df1['class']

```

In [61]: `X`

Out[61]:

	cap- shape	cap- surface	cap- color	bruises	odor	gill- attachment	gill- spacing	gill- size	gill- color	stalk- shape	...	stalk- surface- below- ring	stalk- color- above- ring	stalk- color- below- ring	veil- type	veil- color	ring- number	ring- type	spore- print- color	I
7701	2	2	4	0	5	0	0	0	11	0	...	2	5	5	0	1	1	4	0	
1673	5	2	8	0	5	1	1	0	5	1	...	0	7	7	0	2	1	0	3	
1567	5	2	8	0	5	1	1	0	3	1	...	2	7	7	0	2	1	0	3	
7197	3	2	4	0	7	1	0	1	0	1	...	1	6	7	0	2	1	0	7	
3182	5	0	8	0	1	1	0	1	5	0	...	2	7	7	0	2	1	4	2	
...	
7649	0	2	8	0	5	1	1	0	7	0	...	2	7	7	0	2	2	4	7	
1722	5	0	4	0	5	1	1	0	3	1	...	0	7	7	0	2	1	0	3	
4445	2	0	3	0	2	1	0	0	2	0	...	1	0	4	0	2	1	2	1	

	cap- shape	cap- surface	cap- color	bruises	odor	gill- attachment	gill- spacing	gill- size	gill- color	stalk- shape	...	stalk- surface- below- ring	stalk- color- above- ring	stalk- color- below- ring	veil- type	veil- color	ring- number	ring- type	spore- print- color	
1431	5	0	4	1	5	1	0	0	7	1	...	2	7	7	0	2	1	4	2	
2463	5	3	3	1	5	1	0	0	9	1	...	2	7	6	0	2	1	4	2	

8124 rows × 22 columns



In [62]:

```
Y
```

Out[62]:

```
7701    0
1673    0
1567    0
7197    1
3182    1
..
7649    0
1722    0
4445    1
1431    0
2463    0
Name: class, Length: 8124, dtype: int32
```

In [63]:

```
#checking for null values
df1.isna().sum()
```

Out[63]:

```
class                0
cap-shape            0
cap-surface          0
cap-color            0
bruises              0
odor                 0
gill-attachment      0
gill-spacing         0
gill-size            0
gill-color           0
stalk-shape          0
stalk-root           0
stalk-surface-above-ring 0
stalk-surface-below-ring 0
stalk-color-above-ring 0
```

```
stalk-color-below-ring      0
veil-type                   0
veil-color                  0
ring-number                 0
ring-type                   0
spore-print-color           0
population                  0
habitat                     0
dtype: int64
```

```
In [64]: #checking for duplicates
df1.duplicated().sum()
```

Out[64]: 0

```
In [65]: #checking for unique value counts
for i in df1.columns:
    print(df1[str(i)].value_counts())
    print('\n')
```

```
0    4208
1    3916
Name: class, dtype: int64
```

```
5    3656
2    3152
3     828
0     452
4      32
1       4
Name: cap-shape, dtype: int64
```

```
3    3244
2    2556
0    2320
1       4
Name: cap-surface, dtype: int64
```

```
4    2284
3    1840
2    1500
9    1072
8     1040
```

0	168
5	144
1	44
7	16
6	16

Name: cap-color, dtype: int64

0	4748
1	3376

Name: bruises, dtype: int64

5	3528
2	2160
7	576
8	576
3	400
0	400
6	256
1	192
4	36

Name: odor, dtype: int64

1	7914
0	210

Name: gill-attachment, dtype: int64

0	6812
1	1312

Name: gill-spacing, dtype: int64

0	5612
1	2512

Name: gill-size, dtype: int64

0	1728
7	1492
10	1202
5	1048
2	752
3	732
9	492

4	408
1	96
11	86
6	64
8	24

Name: gill-color, dtype: int64

1	4608
0	3516

Name: stalk-shape, dtype: int64

1	3776
0	2480
3	1120
2	556
4	192

Name: stalk-root, dtype: int64

2	5176
1	2372
0	552
3	24

Name: stalk-surface-above-ring, dtype: int64

2	4936
1	2304
0	600
3	284

Name: stalk-surface-below-ring, dtype: int64

7	4464
6	1872
3	576
4	448
0	432
5	192
2	96
1	36
8	8

Name: stalk-color-above-ring, dtype: int64

7	4384
6	1872
3	576
4	512
0	432
5	192
2	96
1	36
8	24

Name: stalk-color-below-ring, dtype: int64

0	8124
---	------

Name: veil-type, dtype: int64

2	7924
1	96
0	96
3	8

Name: veil-color, dtype: int64

1	7488
2	600
0	36

Name: ring-number, dtype: int64

4	3968
0	2776
2	1296
1	48
3	36

Name: ring-type, dtype: int64

7	2388
3	1968
2	1872
1	1632
5	72
6	48
8	48
4	48
0	48

Name: spore-print-color, dtype: int64

```
4    4040
5    1712
3    1248
2     400
0     384
1     340
Name: population, dtype: int64
```

```
0    3148
1    2148
4    1144
2     832
5     368
3     292
6     192
Name: habitat, dtype: int64
```

```
In [66]: X_train, X_test, Y_train, Y_test = train_test_split(X, Y, test_size = 0.3)
```

```
In [67]: model = DecisionTreeClassifier(criterion = 'entropy', max_depth = 1)
```

```
In [68]: AdaBoost = AdaBoostClassifier(base_estimator = model, n_estimators = 400, learning_rate = 1)
```

```
In [69]: boost_model = AdaBoost.fit(X_train, Y_train)
```

```
In [70]: y_pred = boost_model.predict(X_test)
```

```
In [71]: predictions = metrics.accuracy_score(Y_test, y_pred)
```

```
In [72]: print('The Accuracy is: ', predictions * 100, '%')
```

```
The Accuracy is:  100.0 %
```

```
In [ ]:
```

```
In [ ]:
```