# ENSEMBLES DEMO PROGRAM - 13NOV2021

In [1]:
```
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
```

In [4]:
```
df1 = pd.read_csv('C:/Users/USER/Desktop/MLTHEORY-10NOV2021/MLENSEMBLESDOCS-04NOV2021/mushroomdataset/mushrooms.csv')
```

In [12]:
```
df2 = df1.sample(frac = 1)
```

In [13]:
```
df2.shape
```

Out[13]:
```
(8124, 23)
```

In [14]:
```
df3 = df1.sample(frac = 0.5)
```

In [16]:
```
df3.shape
```

Out[16]:
```
(4062, 23)
```

In [17]:
```
df1.columns
```

Out[17]:
```
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 [19]:
```
for label in df1.columns:
    df1[label] = LabelEncoder().fit(df1[label]).transform(df1[label])
```

In [22]:
```
df1.info()
```

```
<class 'pandas.core.frame.DataFrame'>
```

```
Int64Index: 8124 entries, 6234 to 3414
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 [26]:
```
df1.describe()
```

Out[26]:

| | class | cap-shape | cap-surface | cap-color | bruises | odor | gill-attachment | gill-spacing | gill-size | gill-color | ... | stalk-surface-below-ring |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **count** | 8124.000000 | 8124.000000 | 8124.000000 | 8124.000000 | 8124.000000 | 8124.000000 | 8124.000000 | 8124.000000 | 8124.000000 | 8124.000000 | ... | 8124.000000 |
| **mean** | 0.482029 | 3.348104 | 1.827671 | 4.504677 | 0.415559 | 4.144756 | 0.974151 | 0.161497 | 0.309207 | 4.810684 | ... | 1.603644 |
| **std** | 0.499708 | 1.604329 | 1.229873 | 2.545821 | 0.492848 | 2.103729 | 0.158695 | 0.368011 | 0.462195 | 3.540359 | ... | 0.675974 |
| **min** | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | ... | 0.000000 |
| **25%** | 0.000000 | 2.000000 | 0.000000 | 3.000000 | 0.000000 | 2.000000 | 1.000000 | 0.000000 | 0.000000 | 2.000000 | ... | 1.000000 |
| **50%** | 0.000000 | 3.000000 | 2.000000 | 4.000000 | 0.000000 | 5.000000 | 1.000000 | 0.000000 | 0.000000 | 5.000000 | ... | 2.000000 |

| | class | cap-shape | cap-surface | cap-color | bruises | odor | gill-attachment | gill-spacing | gill-size | gill-color | ... | stalk-surface-below-ring |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **75%** | 1.000000 | 5.000000 | 3.000000 | 8.000000 | 1.000000 | 5.000000 | 1.000000 | 0.000000 | 1.000000 | 7.000000 | ... | 2.000000 |
| **max** | 1.000000 | 5.000000 | 3.000000 | 9.000000 | 1.000000 | 8.000000 | 1.000000 | 1.000000 | 1.000000 | 11.000000 | ... | 3.000000 |

8 rows × 23 columns

In [27]:
```
df1.isna().sum()
```

Out[27]:
```
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 [28]:
```
X = df1.drop(['class'], axis = 1)
Y = df1['class']
```

In [33]:
```
X.shape
```

Out[33]:
```
(8124, 22)
```

In [31]:
```
Y
```

Out[31]:
```
6234    1
3381    1
112     0
2273    0
3977    1
       ..
6786    1
7543    0
1957    0
6191    1
3414    0
Name: class, Length: 8124, dtype: int32
```

In [34]:
```
Y.shape
```

Out[34]:
```
(8124,)
```

In [40]:
```
X_train, X_test, Y_train, Y_test = train_test_split(X, Y, test_size = 0.3)
```

In [41]:
```
X_train.shape
```

Out[41]:
```
(5686, 22)
```

In [42]:
```
Y_train.shape
```

Out[42]:
```
(5686,)
```

In [44]:
```
X_test.shape
```

Out[44]:
```
(2438, 22)
```

In [45]:
```
Y_test.shape
```

Out[45]:
```
(2438,)
```

In [ ]:
```

```