```python
import os #paths to file
import numpy as np # linear algebra
import pandas as pd # data processing
import warnings# warning filter

#ploting libraries
import matplotlib.pyplot as plt
import seaborn as sns

#feature engineering
from sklearn.preprocessing import OneHotEncoder
from sklearn.preprocessing import LabelEncoder

#train test split
from sklearn.model_selection import train_test_split

#metrics
from sklearn.metrics import f1_score
from sklearn.metrics import accuracy_score
from sklearn.metrics import recall_score
from sklearn.metrics import precision_score
from sklearn.model_selection import cross_validate

#cross validation
from sklearn.model_selection  import cross_val_score as CVS


#ML models
from xgboost import XGBClassifier
from sklearn.tree import DecisionTreeClassifier
from sklearn.ensemble import RandomForestClassifier
from sklearn.linear_model import LogisticRegression
from sklearn.svm import SVC

#default theme and settings
sns.set(context='notebook', style='darkgrid', palette='deep', font='sans-serif
', font_scale=1, color_codes=False, rc=None)
pd.set_option('display.max_columns', None)

#warning handle
warnings.filterwarnings("always")
warnings.filterwarnings("ignore")


#list all files under the input directory
for dirname, _, filenames in os.walk('/kaggle/input'):
    for filename in filenames:
        print(os.path.join(dirname, filename))


#path for the dataset
Mushroom_path = "/kaggle/input/mushroom-classification/mushrooms.csv"
Mushroom_df = pd.read_csv(Mushroom_path)
Mushroom_df.head()
```

```python
Mushroom_df = pd.read_csv(Mushroom_path)
Mushroom_df.head()
```

	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
0	p	x	s	n	t	p	f	c	n	k	e	e	s	s	w	w	p	w	o	p	k	s	u
1	e	x	s	y	t	a	f	c	b	k	e	c	s	s	w	w	p	w	o	p	n	n	g
2	e	b	s	w	t	l	f	c	b	n	e	c	s	s	w	w	p	w	o	p	n	n	m
3	p	x	y	w	t	p	f	c	n	n	e	e	s	s	w	w	p	w	o	p	k	s	u
4	e	x	s	g	f	n	f	w	b	k	t	e	s	s	w	w	p	w	o	e	n	a	g

Shape

In [5]:
```python
"Data Shape (row, col): {}".format(Mushroom_df.shape)
```

Out[5]:
```
'Data Shape (row, col): (8124, 23)'
```

Information

In [6]:
```python
Mushroom_df.info()
```
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8124 entries, 0 to 8123
Data columns (total 23 columns):
 #   Column                    Non-Null Count  Dtype
---  ------                    --------------  -----
 0   class                     8124 non-null   object
 1   cap-shape                 8124 non-null   object
```

```
 2   cap-surface               8124 non-null   object
 3   cap-color                 8124 non-null   object
 4   bruises                   8124 non-null   object
 5   odor                      8124 non-null   object
 6   gill-attachment           8124 non-null   object
 7   gill-spacing              8124 non-null   object
 8   gill-size                 8124 non-null   object
 9   gill-color                8124 non-null   object
 10  stalk-shape               8124 non-null   object
 11  stalk-root                8124 non-null   object
 12  stalk-surface-above-ring  8124 non-null   object
 13  stalk-surface-below-ring  8124 non-null   object
 14  stalk-color-above-ring    8124 non-null   object
 15  stalk-color-below-ring    8124 non-null   object
 16  veil-type                 8124 non-null   object
 17  veil-color                8124 non-null   object
 18  ring-number               8124 non-null   object
 19  ring-type                 8124 non-null   object
 20  spore-print-color         8124 non-null   object
 21  population                8124 non-null   object
 22  habitat                   8124 non-null   object
dtypes: object(23)
memory usage: 1.4+ MB
```

Missing values 🚫

We have no missing values! let's confirm it:

```
In [7]:
#missing values
Mushroom_df.isnull().sum()

Out[7]:
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 [8]:
M_cols = Mushroom_df.columns.to_list()
print("Value Distribution:\n")
for col in M_cols:
    print(col,"\n",Mushroom_df[col].value_counts(),"\n\n")
```

Value Distribution:

class
 e   4208
p   3916
Name: class, dtype: int64

cap-shape
 x   3656
f   3152
k   828
b   452
s   32
c   4
Name: cap-shape, dtype: int64

cap-surface
 y   3244
s   2556
f   2320
g   4
Name: cap-surface, dtype: int64

cap-color
 n   2284
g   1840
e   1500
y   1072
w   1040
b   168
p   144
c   44
u   16
r   16
Name: cap-color, dtype: int64

bruises
 f   4748
t   3376
Name: bruises, dtype: int64

odor
 n   3528
f   2160
s   576
y   576
```

```
l     400
a     400
p     256
c     192
m      36
Name: odor, dtype: int64


gill-attachment
 f    7914
a     210
Name: gill-attachment, dtype: int64


gill-spacing
 c    6812
w    1312
Name: gill-spacing, dtype: int64


gill-size
 b    5612
n    2512
Name: gill-size, dtype: int64


gill-color
 b    1728
p    1492
w    1202
n    1048
g     752
h     732
u     492
k     408
e      96
y      86
o      64
r      24
Name: gill-color, dtype: int64


stalk-shape
 t    4608
e    3516
Name: stalk-shape, dtype: int64


stalk-root
 b    3776
?    2480
e    1120
c     556
r     192
Name: stalk-root, dtype: int64
```

```
stalk-surface-above-ring
 s    5176
k    2372
f     552
y      24
Name: stalk-surface-above-ring, dtype: int64


stalk-surface-below-ring
 s    4936
k    2304
f     600
y     284
Name: stalk-surface-below-ring, dtype: int64


stalk-color-above-ring
 w    4464
p    1872
g     576
n     448
b     432
o     192
e      96
c      36
y       8
Name: stalk-color-above-ring, dtype: int64


stalk-color-below-ring
 w    4384
p    1872
g     576
n     512
b     432
o     192
e      96
c      36
y      24
Name: stalk-color-below-ring, dtype: int64


veil-type
 p    8124
Name: veil-type, dtype: int64


veil-color
 w    7924
o      96
n      96
y       8
Name: veil-color, dtype: int64
```

```
ring-number
 o    7488
 t     600
 n      36
Name: ring-number, dtype: int64


ring-type
 p    3968
 e    2776
 l    1296
 f      48
 n      36
Name: ring-type, dtype: int64


spore-print-color
 w    2388
 n    1968
 k    1872
 h    1632
 r      72
 b      48
 o      48
 u      48
 y      48
Name: spore-print-color, dtype: int64


population
 v    4040
 y    1712
 s    1248
 n     400
 a     384
 c     340
Name: population, dtype: int64


habitat
 d    3148
 g    2148
 p    1144
 l     832
 u     368
 m     292
 w     192
Name: habitat, dtype: int64
```

arget plot

We will make a detailed plot for our target to get a clear idea of it's distribution:

In [9]:

```python
total = float(len(Mushroom_df[M_cols[0]]))
plt.figure(figsize=(6,6))
sns.set(style="darkgrid")
ax = sns.countplot(Mushroom_df[M_cols[0]])
for p in ax.patches:
    height = p.get_height()
    ax.text(p.get_x()+p.get_width()/2.,height + 3,'{:1.2f}'.format(height/total),ha="center")
plt.title("Target Plot", fontsize = 20)
plt.show()
```

Unvariate plots

In [10]:

```python
for col in M_cols[1:]:
    plt.figure(figsize=(10,4))
    sns.countplot(x=col , data=Mushroom_df ,palette='icefire')
    plt.title(col, fontsize=14)
    plt.show()
    print("% of total:")
    print(round((Mushroom_df[col].value_counts()/Mushroom_df.shape[0]),4)*100)
```

```
% of total:
x    45.00
f    38.80
k    10.19
b     5.56
s     0.39
c     0.05
Name: cap-shape, dtype: float64

% of total:
y    39.93
s    31.46
f    28.56
g     0.05
Name: cap-surface, dtype: float64

% of total:
n    28.11
g    22.65
e    18.46
y    13.20
w    12.80
b     2.07
p     1.77
c     0.54
u     0.20
r     0.20
Name: cap-color, dtype: float64

% of total:
f    58.44
t    41.56
```

Name: bruises, dtype: float64

% of total:
```
n    43.43
f    26.59
s     7.09
y     7.09
l     4.92
a     4.92
p     3.15
c     2.36
m     0.44
```
Name: odor, dtype: float64

% of total:
```
f    97.42
a     2.58
```
Name: gill-attachment, dtype: float64

% of total:
```
c    83.85
w    16.15
```
Name: gill-spacing, dtype: float64

% of total:
```
b    69.08
n    30.92
```
Name: gill-size, dtype: float64

% of total:
```
b    21.27
p    18.37
w    14.80
n    12.90
g     9.26
h     9.01
u     6.06
k     5.02
e     1.18
y     1.06
o     0.79
r     0.30
```
Name: gill-color, dtype: float64

% of total:
```
t    56.72
e    43.28
```
Name: stalk-shape, dtype: float64

% of total:
```
b    46.48
?    30.53
e    13.79
c     6.84
r     2.36
```

```
Name: stalk-root, dtype: float64

% of total:
s    63.71
k    29.20
f     6.79
y     0.30
Name: stalk-surface-above-ring, dtype: float64

% of total:
s    60.76
k    28.36
f     7.39
y     3.50
Name: stalk-surface-below-ring, dtype: float64

% of total:
w    54.95
p    23.04
g     7.09
n     5.51
b     5.32
o     2.36
e     1.18
c     0.44
y     0.10
Name: stalk-color-above-ring, dtype: float64

% of total:
w    53.96
p    23.04
g     7.09
n     6.30
b     5.32
o     2.36
e     1.18
c     0.44
y     0.30
Name: stalk-color-below-ring, dtype: float64

% of total:
p    100.0
Name: veil-type, dtype: float64

% of total:
w    97.54
o     1.18
n     1.18
y     0.10
Name: veil-color, dtype: float64

% of total:
o    92.17
t     7.39
n     0.44
```

Name: ring-number, dtype: float64

% of total:
p   48.84
e   34.17
l   15.95
f   0.59
n   0.44
Name: ring-type, dtype: float64

% of total:
w   29.39
n   24.22
k   23.04
h   20.09
r   0.89
b   0.59
o   0.59
u   0.59
y   0.59
Name: spore-print-color, dtype: float64

% of total:
v   49.73
y   21.07
s   15.36
n   4.92
a   4.73
c   4.19
Name: population, dtype: float64

% of total:
d   38.75
g   26.44
p   14.08
l   10.24
u   4.53
m   3.59
w   2.36
Name: habitat, dtype: float64