Context

Although this dataset was originally contributed to the UCI Machine Learning repository nearly 30 years ago, mushroom hunting (otherwise known as "shrooming") is enjoying new peaks in popularity. Learn which features spell certain death and which are most palatable in this dataset of mushroom characteristics. And how certain can your model be?

Content

This dataset includes descriptions of hypothetical samples corresponding to 23 species of gilled mushrooms in the Agaricus and Lepiota Family Mushroom drawn from The Audubon Society Field Guide to North American Mushrooms (1981). 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.

Time period: Donated to UCI ML 27 April 1987

Inspiration

What types of machine learning models perform best on this dataset?

Which features are most indicative of a poisonous mushroom?


```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
```

```
!pip install dataprep
```

```
Requirement already satisfied: dataprep in /usr/local/lib/python3.7/dist-packages (0
Requirement already satisfied: pydantic<2.0,>=1.6 in /usr/local/lib/python3.7/dist-p
Requirement already satisfied: dask[array,dataframe,delayed]<3.0,>=2.25 in /usr/loca
Requirement already satisfied: numpy<2,>=1 in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: nltk<4.0,>=3.5 in /usr/local/lib/python3.7/dist-packa
Requirement already satisfied: regex<2021.0.0,>=2020.10.15 in /usr/local/lib/python3
Requirement already satisfied: scipy<2,>=1 in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: bokeh<3,>=2 in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: jsonpath-ng<2.0,>=1.5 in /usr/local/lib/python3.7/dis
Requirement already satisfied: tqdm<5.0,>=4.48 in /usr/local/lib/python3.7/dist-pack
Requirement already satisfied: pandas<2,>=1 in /usr/local/lib/python3.7/dist-package
```

```
Requirement already satisfied: wordcloud<2.0,>=1.8 in /usr/local/lib/python3.7/dist-
Requirement already satisfied: jinja2<3.0,>=2.11 in /usr/local/lib/python3.7/dist-pa
Requirement already satisfied: bottleneck<2.0,>=1.3 in /usr/local/lib/python3.7/dist
Requirement already satisfied: aiohttp<4.0,>=3.6 in /usr/local/lib/python3.7/dist-pa
Requirement already satisfied: ipywidgets<8.0,>=7.5 in /usr/local/lib/python3.7/dist
Requirement already satisfied: yarl<2.0,>=1.0 in /usr/local/lib/python3.7/dist-packa
Requirement already satisfied: typing-extensions>=3.6.5 in /usr/local/lib/python3.7/
Requirement already satisfied: attrs>=17.3.0 in /usr/local/lib/python3.7/dist-packag
Requirement already satisfied: async-timeout<4.0,>=3.0 in /usr/local/lib/python3.7/d
Requirement already satisfied: chardet<5.0,>=2.0 in /usr/local/lib/python3.7/dist-pa
Requirement already satisfied: multidict<7.0,>=4.5 in /usr/local/lib/python3.7/dist-
Requirement already satisfied: tornado>=5.1 in /usr/local/lib/python3.7/dist-package
Requirement already satisfied: PyYAML>=3.10 in /usr/local/lib/python3.7/dist-package
Requirement already satisfied: pillow>=7.1.0 in /usr/local/lib/python3.7/dist-packag
Requirement already satisfied: packaging>=16.8 in /usr/local/lib/python3.7/dist-pack
Requirement already satisfied: python-dateutil>=2.1 in /usr/local/lib/python3.7/dist
Requirement already satisfied: toolz>=0.8.2 in /usr/local/lib/python3.7/dist-package
Requirement already satisfied: cloudpickle>=0.2.2 in /usr/local/lib/python3.7/dist-p
Requirement already satisfied: partd>=0.3.10 in /usr/local/lib/python3.7/dist-packag
Requirement already satisfied: fsspec>=0.6.0 in /usr/local/lib/python3.7/dist-packag
Requirement already satisfied: traitlets>=4.3.1 in /usr/local/lib/python3.7/dist-pac
Requirement already satisfied: ipykernel>=4.5.1 in /usr/local/lib/python3.7/dist-pac
Requirement already satisfied: jupyterlab-widgets>=1.0.0 in /usr/local/lib/python3.7
Requirement already satisfied: widgetsnbextension~=3.5.0 in /usr/local/lib/python3.7
Requirement already satisfied: nbformat>=4.2.0 in /usr/local/lib/python3.7/dist-pack
Requirement already satisfied: ipython>=4.0.0 in /usr/local/lib/python3.7/dist-packa
Requirement already satisfied: jupyter-client in /usr/local/lib/python3.7/dist-packa
Requirement already satisfied: pexpect in /usr/local/lib/python3.7/dist-packages (fr
Requirement already satisfied: pickleshare in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: pygments in /usr/local/lib/python3.7/dist-packages (f
Requirement already satisfied: prompt-toolkit<2.0.0,>=1.0.4 in /usr/local/lib/python
Requirement already satisfied: simplegeneric>0.8 in /usr/local/lib/python3.7/dist-pa
Requirement already satisfied: setuptools>=18.5 in /usr/local/lib/python3.7/dist-pac
Requirement already satisfied: decorator in /usr/local/lib/python3.7/dist-packages (
Requirement already satisfied: MarkupSafe>=0.23 in /usr/local/lib/python3.7/dist-pac
Requirement already satisfied: ply in /usr/local/lib/python3.7/dist-packages (from j
Requirement already satisfied: six in /usr/local/lib/python3.7/dist-packages (from j
Requirement already satisfied: ipython-genutils in /usr/local/lib/python3.7/dist-pac
Requirement already satisfied: jupyter-core in /usr/local/lib/python3.7/dist-package
Requirement already satisfied: jsonschema!=2.5.0,>=2.4 in /usr/local/lib/python3.7/d
Requirement already satisfied: click in /usr/local/lib/python3.7/dist-packages (from
Requirement already satisfied: joblib in /usr/local/lib/python3.7/dist-packages (fro
Requirement already satisfied: pyparsing>=2.0.2 in /usr/local/lib/python3.7/dist-pac
Requirement already satisfied: pytz>=2017.2 in /usr/local/lib/python3.7/dist-package
Requirement already satisfied: locket in /usr/local/lib/python3.7/dist-packages (fro
Requirement already satisfied: wcwidth in /usr/local/lib/python3.7/dist-packages (fr
Requirement already satisfied: notebook>=4.4.1 in /usr/local/lib/python3.7/dist-pack
```

Saving...

```
from dataprep.eda import create_report,plot,plot_correlation,plot_missing
```

```
df = pd.read_csv("/content/drive/MyDrive/Colab Notebooks/kaggle/mushrooms.csv")
```

```
df.head()
```

	class	cap-shape	cap-surface	cap-color	bruises	odor	gill-attachment	gill-spacing	gill-size	gill-color	stall shaj
0	p	x	s	n	t	p	f	c	n	k	
1	e	x	s	y	t	a	f	c	b	k	
2	e	b	s	w	t	l	f	c	b	n	
3	p	x	y	w	t	p	f	c	n	n	
4	e	x	s	g	f	n	f	w	b	k	

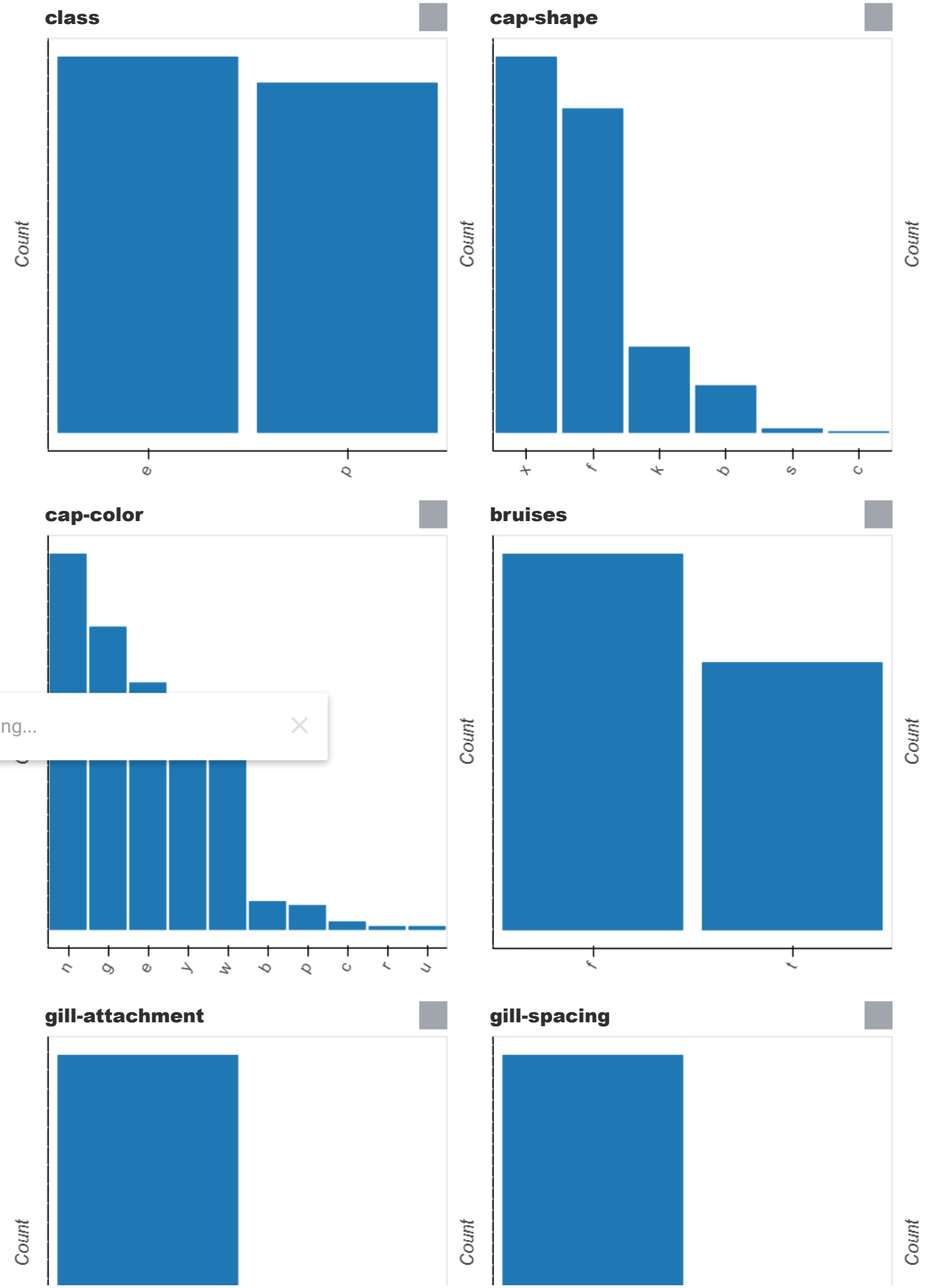
```
create_report(df,title = "My Report")
```

```
create_report(df).show_browser()
```

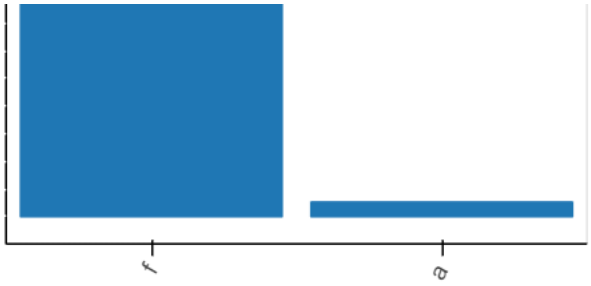
```
plot(df)
```

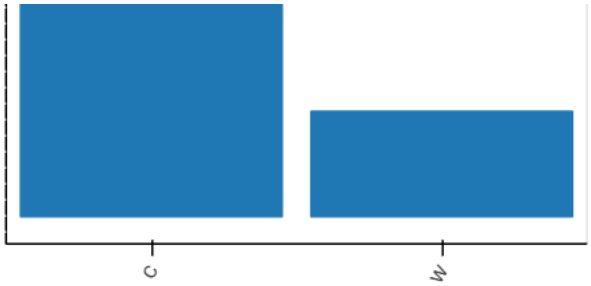
Saving...

Show Stats and Insights

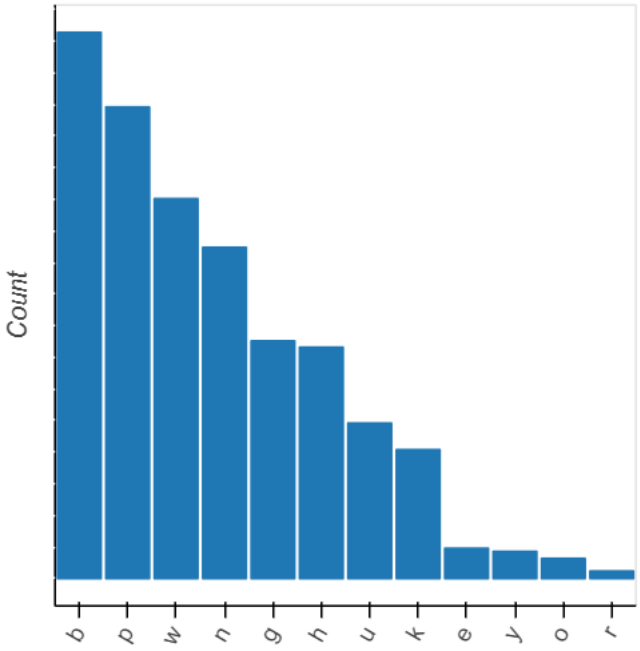


Saving...

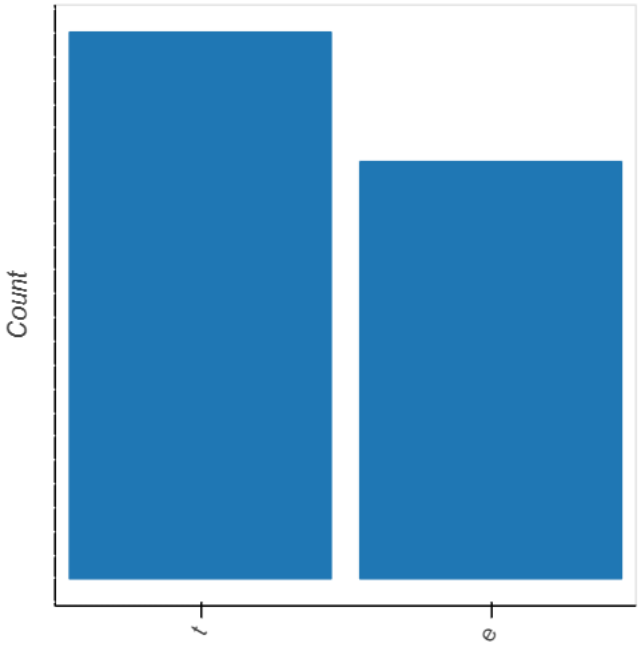




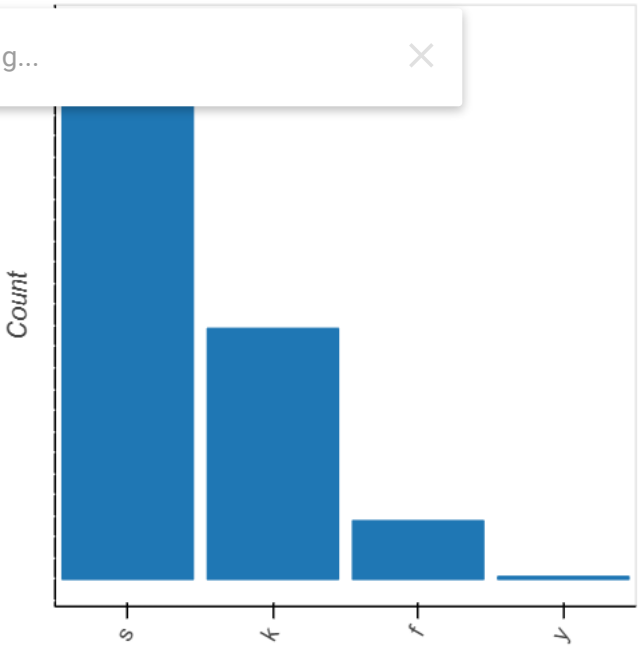
gill-color



stalk-shape

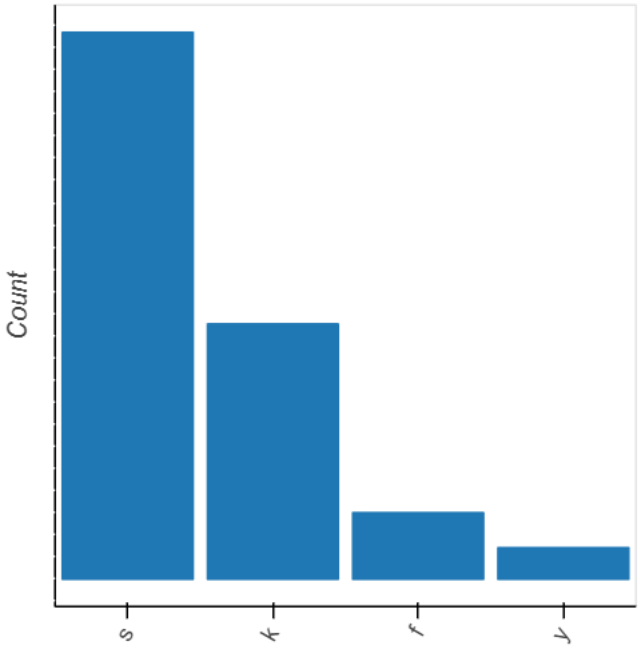


stalk-surface-above-ring

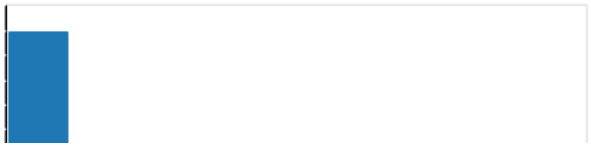


Saving...

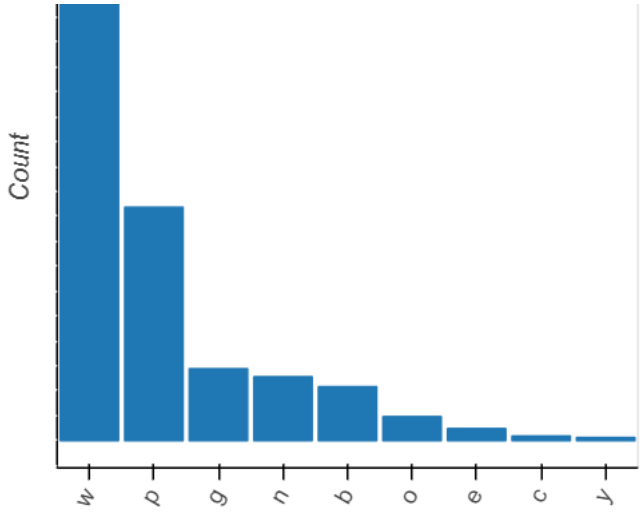
stalk-surface-below-ring

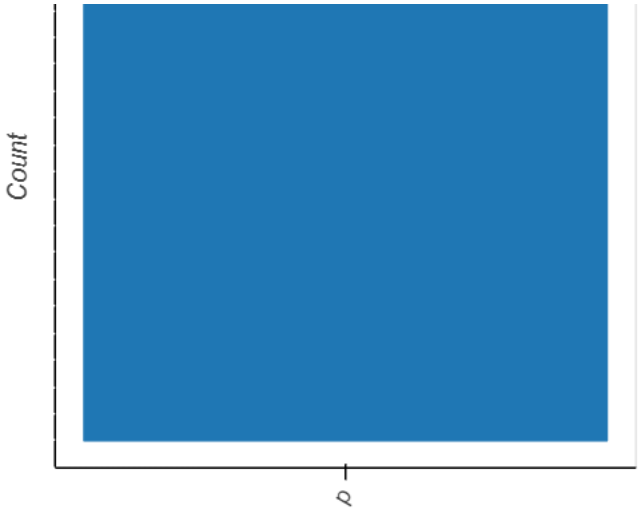


stalk-color-below-ring

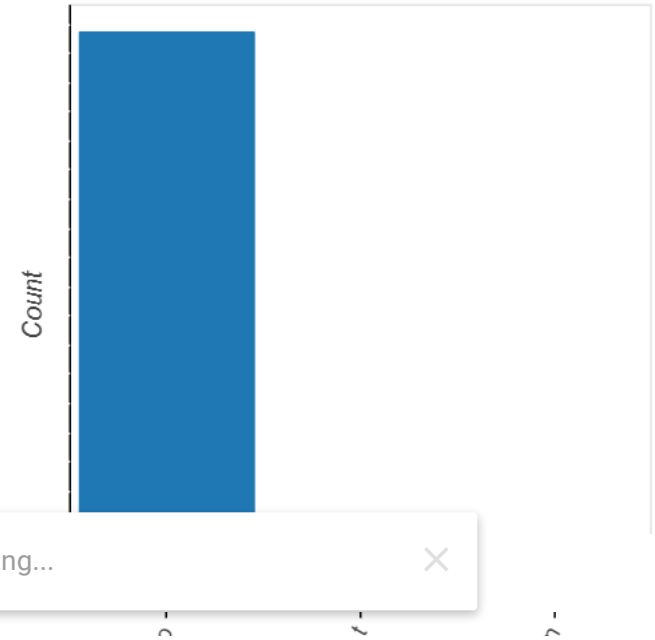


veil-type

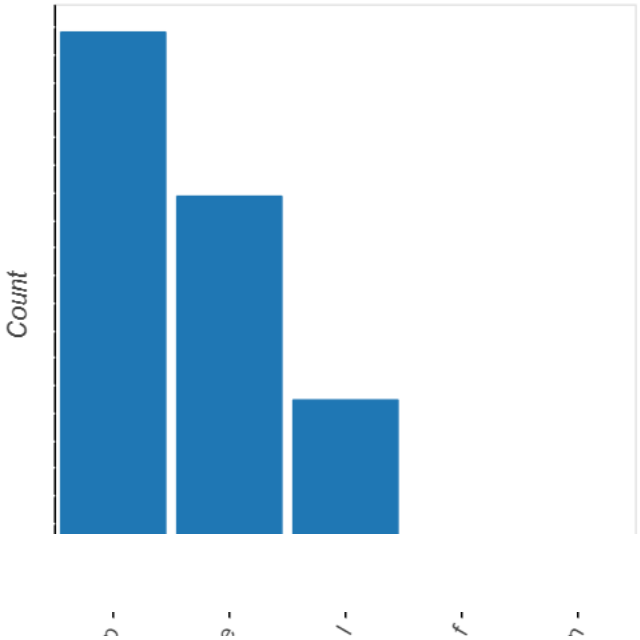




ring-number



ring-type



Saving...

Independent And Dependent Features

```
X = df.drop("class",axis=1)
```

```
y=df["class"]
```

```
X.head()
```

	cap-shape	cap-surface	cap-color	bruises	odor	gill-attachment	gill-spacing	gill-size	gill-color	stalk-shape	sta…
0	x	s	n	t	p	f	c	n	k	e	

```
y.value_counts(normalize=True)
```

```
e    0.517971
p    0.482029
Name: class, dtype: float64
```

4	x	s	g	f	n	f	w	b	k	t	

```
!python setup.py install
!python -m pip install --upgrade pip setuptools wheel
!pip install evalml
```

```
python3: can't open file 'setup.py': [Errno 2] No such file or directory
Requirement already satisfied: pip in /usr/local/lib/python3.7/dist-packages (21.1.1
Requirement already satisfied: setuptools in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: wheel in /usr/local/lib/python3.7/dist-packages (0.36
WARNING: Running pip as root will break packages and permissions. You should install
Requirement already satisfied: evalml in /usr/local/lib/python3.7/dist-packages (0.2
Requirement already satisfied: sktime>=0.5.3 in /usr/local/lib/python3.7/dist-packag
Requirement already satisfied: matplotlib>=3.3.3 in /usr/local/lib/python3.7/dist-pa
Requirement already satisfied: cloudpickle>=0.2.2 in /usr/local/lib/python3.7/dist-p
Requirement already satisfied: pyzmq<22.0.0 in /usr/local/lib/python3.7/dist-package
Requirement already satisfied: nlp-primitives>=1.1.0 in /usr/local/lib/python3.7/dis
Requirement already satisfied: statsmodels>=0.12.2 in /usr/local/lib/python3.7/dist-
Requirement already satisfied: category-encoders>=2.0.0 in /usr/local/lib/python3.7/
Requirement already satisfied: dask>=2.12.0 in /usr/local/lib/python3.7/dist-package
Requirement already satisfied: scikit-optimize>=0.8.1 in /usr/local/lib/python3.7/di
Requirement already satisfied: networkx>=2.5 in /usr/local/lib/python3.7/dist-packag
Requirement already satisfied: scikit-learn>=0.23.1 in /usr/local/lib/python3.7/dist
Requirement already satisfied: plotly>=4.14.0 in /usr/local/lib/python3.7/dist-packa
Requirement already satisfied: woodwork==0.0.11 in /usr/local/lib/python3.7/dist-pac
Requirement already satisfied: click>=7.0.0 in /usr/local/lib/python3.7/dist-package
Requirement already satisfied: ipywidgets>=7.5 in /usr/local/lib/python3.7/dist-pack
Requirement already satisfied: imbalanced-learn>=0.8.0 in /usr/local/lib/python3.7/d
Requirement already satisfied: numpy>=1.19.1 in /usr/local/lib/python3.7/dist-packag
Requirement already satisfied: pandas>=1.1.0 in /usr/local/lib/python3.7/dist-packag
Requirement already satisfied: colorama in /usr/local/lib/python3.7/dist-packages (f
Requirement already satisfied: shap>=0.36.0 in /usr/local/lib/python3.7/dist-package
Requirement already satisfied: texttable>=1.6.2 in /usr/local/lib/python3.7/dist-pac
Requirement already satisfied: featuretools>=0.20.0 in /usr/local/lib/python3.7/dist
Requirement already satisfied: kaleido>=0.1.0 in /usr/local/lib/python3.7/dist-packa
Requirement already satisfied: seaborn>=0.11.1 in /usr/local/lib/python3.7/dist-pack
Requirement already satisfied: graphviz>=0.13 in /usr/local/lib/python3.7/dist-packa
Requirement already satisfied: xgboost<1.3.0,>=0.82 in /usr/local/lib/python3.7/dist
Requirement already satisfied: lightgbm<3.1.0,>=2.3.1 in /usr/local/lib/python3.7/di
Requirement already satisfied: requirements-parser>=0.2.0 in /usr/local/lib/python3.
Requirement already satisfied: catboost>=0.20 in /usr/local/lib/python3.7/dist-packa
Requirement already satisfied: scipy>=1.2.1 in /usr/local/lib/python3.7/dist-package
Requirement already satisfied: psutil>=5.6.3 in /usr/local/lib/python3.7/dist-packag
Requirement already satisfied: six in /usr/local/lib/python3.7/dist-packages (from c
Requirement already satisfied: patsy>=0.5.1 in /usr/local/lib/python3.7/dist-package
```

Saving...

```
Requirement already satisfied: pyyaml in /usr/local/lib/python3.7/dist-packages (frc
Requirement already satisfied: tqdm>=4.32.0 in /usr/local/lib/python3.7/dist-package
Requirement already satisfied: distributed>=2.12.0 in /usr/local/lib/python3.7/dist-
Requirement already satisfied: fsspec>=0.6.0 in /usr/local/lib/python3.7/dist-packag
Requirement already satisfied: partd>=0.3.10 in /usr/local/lib/python3.7/dist-packag
Requirement already satisfied: toolz>=0.8.2 in /usr/local/lib/python3.7/dist-package
Requirement already satisfied: msgpack>=0.6.0 in /usr/local/lib/python3.7/dist-packa
Requirement already satisfied: sortedcontainers!=2.0.0,!=2.0.1 in /usr/local/lib/pyt
Requirement already satisfied: tblib>=1.6.0 in /usr/local/lib/python3.7/dist-package
Requirement already satisfied: setuptools in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: tornado>=5 in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: zict>=0.1.3 in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: joblib>=0.11 in /usr/local/lib/python3.7/dist-package
Requirement already satisfied: jupyterlab-widgets>=1.0.0 in /usr/local/lib/python3.7
Requirement already satisfied: ipykernel>=4.5.1 in /usr/local/lib/python3.7/dist-pac
Requirement already satisfied: nbformat>=4.2.0 in /usr/local/lib/python3.7/dist-pack
Requirement already satisfied: traitlets>=4.3.1 in /usr/local/lib/python3.7/dist-pac
Requirement already satisfied: widgetsnbextension~=3.5.0 in /usr/local/lib/python3.7
Requirement already satisfied: ipython>=4.0.0 in /usr/local/lib/python3.7/dist-packa
```

```
import evalml
from evalml import AutoMLSearch
```

```
### Train Test Split
```

```
X_train,X_test,y_train,y_test = evalml.preprocessing.split_data(X,y,problem_type="binary")
```

Saving...

```
ll_problem_types
```

```
[<ProblemTypes.BINARY: 'binary'>,
 <ProblemTypes.MULTICLASS: 'multiclass'>,
 <ProblemTypes.REGRESSION: 'regression'>,
 <ProblemTypes.TIME_SERIES_REGRESSION: 'time series regression'>,
 <ProblemTypes.TIME_SERIES_BINARY: 'time series binary'>,
 <ProblemTypes.TIME_SERIES_MULTICLASS: 'time series multiclass'>]
```

```
type(X_train)
```

```
woodwork.datatable.DataTable
```

```
X_train.head()
```

Data Column	cap-shape	cap-surface	cap-color	bruises	odor	gill-attachm
Physical Type	category	category	category	category	category	categor
Logical Type	Categorical	Categorical	Categorical	Categorical	Categorical	Categor
Semantic Tag(s)	['category']	['category']	['category']	['category']	['category']	['categ
6373	f	s	n	f	f	
4331	k	y	n	f	n	
6496	f	s	e	f	y	

AutoMLSearch

AutoMLSearch give us most effective and accurate ML model or model to fit datasets

```
from evalml import AutoMLSearch
```

```
automl = AutoMLSearch(X_train = X_train,y_train = y_train,problem_type ="binary",objective='f
```

```
    Generating pipelines to search over...
    Generating pipelines to search over...
    allowed_estimators set to ['Decision Tree Classifier', 'LightGBM Classifier', 'Extra Tr
                               ision Tree Classifier w/ Imputer + One Hot Encoder', 'Lig
                               [ModelFamily.XGBOOST, ModelFamily.DECISION_TREE, ModelFam
```



```
automl.search()
```

```
*****************************

*****************************
* Beginning pipeline search *
* Beginning pipeline search *
*****************************
*****************************


Optimizing for F1.
Optimizing for F1.
Greater score is better.

Greater score is better.

Using SequentialEngine to train and score pipelines.
Using SequentialEngine to train and score pipelines.
Searching up to 1 batches for a total of 9 pipelines.
Searching up to 1 batches for a total of 9 pipelines.
Allowed model families: xgboost, decision_tree, lightgbm, linear_model, extra_trees, ca

Allowed model families: xgboost, decision_tree, lightgbm, linear_model, extra_trees, ca

Evaluating Baseline Pipeline: Mode Baseline Binary Classification Pipeline
Evaluating Baseline Pipeline: Mode Baseline Binary Classification Pipeline
All predicted probabilities have the same value: [0.]
Extreme threshold of 0.0
All predicted probabilities have the same value: [0.]
Extreme threshold of 0.0
All predicted probabilities have the same value: [0.]
Extreme threshold of 0.0
                                    ation Pipeline:
                                    ation Pipeline:
                                    ion
        Starting cross validation
                Training and scoring on fold 0
                        Fold 0: starting training
                        Fold 0: finished training
                        Fold 0: Optimal threshold found (0.000)
                        Fold 0: Scoring trained pipeline
                        Fold 0: F1 score: 0.000
                Training and scoring on fold 1
                        Fold 1: starting training
                        Fold 1: finished training
                        Fold 1: Optimal threshold found (0.000)
                        Fold 1: Scoring trained pipeline
                        Fold 1: F1 score: 0.000
                Training and scoring on fold 2
                        Fold 2: starting training
                        Fold 2: finished training
                        Fold 2: Optimal threshold found (0.000)
                        Fold 2: Scoring trained pipeline
                        Fold 2: F1 score: 0.000
        Finished cross validation - mean F1: 0.000
        Finished cross validation - mean F1: 0.000
```

```
****************************

****************************
* Evaluating Batch Number 1 *
* Evaluating Batch Number 1 *
****************************
****************************


Elastic Net Classifier w/ Imputer + One Hot Encoder + Standard Scaler:
Elastic Net Classifier w/ Imputer + One Hot Encoder + Standard Scaler:
        Starting cross validation
        Starting cross validation
                Training and scoring on fold 0
                        Fold 0: starting training
                        Fold 0: finished training
                        Fold 0: Optimal threshold found (0.533)
                        Fold 0: Scoring trained pipeline
                        Fold 0: F1 score: 0.888
                Training and scoring on fold 1
                        Fold 1: starting training
                        Fold 1: finished training
                        Fold 1: Optimal threshold found (0.535)
                        Fold 1: Scoring trained pipeline
                        Fold 1: F1 score: 0.896
                Training and scoring on fold 2
                        Fold 2: starting training
                        Fold 2: finished training
                        Fold 2: Optimal threshold found (0.535)
                        Fold 2: Scoring trained pipeline
                        Fold 2: F1 score: 0.889
        Finished cross validation - mean F1: 0.891
        Finished cross validation - mean F1: 0.891
```

Saving...

```
Extreme threshold of 1.0
Decision Tree Classifier w/ Imputer + One Hot Encoder:
Decision Tree Classifier w/ Imputer + One Hot Encoder:
        Starting cross validation
        Starting cross validation
                Training and scoring on fold 0
                        Fold 0: starting training
                        Fold 0: finished training
                        Fold 0: Optimal threshold found (0.778)
                        Fold 0: Scoring trained pipeline
                        Fold 0: F1 score: 0.950
                Training and scoring on fold 1
                        Fold 1: starting training
                        Fold 1: finished training
                        Fold 1: Optimal threshold found (0.818)
                        Fold 1: Scoring trained pipeline
                        Fold 1: F1 score: 0.946
                Training and scoring on fold 2
                        Fold 2: starting training
                        Fold 2: finished training
                        Fold 2: Optimal threshold found (0.826)
                        Fold 2: Scoring trained pipeline
                        Fold 2: F1 score: 0.947
```

```
        Finished cross validation - mean F1: 0.948
        Finished cross validation - mean F1: 0.948
Random Forest Classifier w/ Imputer + One Hot Encoder:
Random Forest Classifier w/ Imputer + One Hot Encoder:
        Starting cross validation
        Starting cross validation
                Training and scoring on fold 0
                        Fold 0: starting training
                        Fold 0: finished training
                        Fold 0: Optimal threshold found (0.397)
                        Fold 0: Scoring trained pipeline
                        Fold 0: F1 score: 0.999
                Training and scoring on fold 1
                        Fold 1: starting training
                        Fold 1: finished training
                        Fold 1: Optimal threshold found (0.500)
                        Fold 1: Scoring trained pipeline
                        Fold 1: F1 score: 0.994
                Training and scoring on fold 2
                        Fold 2: starting training
                        Fold 2: finished training
                        Fold 2: Optimal threshold found (0.453)
                        Fold 2: Scoring trained pipeline
                        Fold 2: F1 score: 0.999
        Finished cross validation - mean F1: 0.997
        Finished cross validation - mean F1: 0.997
LightGBM Classifier w/ Imputer + One Hot Encoder:
LightGBM Classifier w/ Imputer + One Hot Encoder:
        Starting cross validation
        Starting cross validation
                Training and scoring on fold 0
                        Fold 0: starting training
                                : finished training
                                : Optimal threshold found (1.000)
                                : Scoring trained pipeline
                        Fold 0: F1 score: 1.000
                Training and scoring on fold 1
                        Fold 1: starting training
                        Fold 1: finished training
                        Fold 1: Optimal threshold found (1.000)
                        Fold 1: Scoring trained pipeline
                        Fold 1: F1 score: 1.000
                Training and scoring on fold 2
                        Fold 2: starting training
                        Fold 2: finished training
                        Fold 2: Optimal threshold found (1.000)
                        Fold 2: Scoring trained pipeline
                        Fold 2: F1 score: 0.998
        Finished cross validation - mean F1: 0.999
        Finished cross validation - mean F1: 0.999
Logistic Regression Classifier w/ Imputer + One Hot Encoder + Standard Scaler:
Logistic Regression Classifier w/ Imputer + One Hot Encoder + Standard Scaler:
        Starting cross validation
        Starting cross validation
                Training and scoring on fold 0
                        Fold 0: starting training
                        Fold 0: finished training
                        Fold 0: Optimal threshold found (0.959)
```

Saving...

```
                        Fold 0: Optimal threshold found (0.939)
                        Fold 0: Scoring trained pipeline
                        Fold 0: F1 score: 1.000
                Training and scoring on fold 1
                        Fold 1: starting training
                        Fold 1: finished training
                        Fold 1: Optimal threshold found (0.983)
                        Fold 1: Scoring trained pipeline
                        Fold 1: F1 score: 0.999
                Training and scoring on fold 2
                        Fold 2: starting training
                        Fold 2: finished training
                        Fold 2: Optimal threshold found (0.986)
                        Fold 2: Scoring trained pipeline
                        Fold 2: F1 score: 0.997
        Finished cross validation - mean F1: 0.998
        Finished cross validation - mean F1: 0.998
XGBoost Classifier w/ Imputer + One Hot Encoder:
XGBoost Classifier w/ Imputer + One Hot Encoder:
        Starting cross validation
        Starting cross validation
                Training and scoring on fold 0
                        Fold 0: starting training
                        Fold 0: finished training
                        Fold 0: Optimal threshold found (0.933)
                        Fold 0: Scoring trained pipeline
                        Fold 0: F1 score: 1.000
                Training and scoring on fold 1
                        Fold 1: starting training
                        Fold 1: finished training
                        Fold 1: Optimal threshold found (0.952)
                        Fold 1: Scoring trained pipeline
                        Fold 1: F1 score: 1.000
                        coring on fold 2
                        : starting training
                        : finished training
                        Fold 2: Optimal threshold found (0.971)
                        Fold 2: Scoring trained pipeline
                        Fold 2: F1 score: 0.996
        Finished cross validation - mean F1: 0.998
        Finished cross validation - mean F1: 0.998
Extra Trees Classifier w/ Imputer + One Hot Encoder:
Extra Trees Classifier w/ Imputer + One Hot Encoder:
        Starting cross validation
        Starting cross validation
                Training and scoring on fold 0
                        Fold 0: starting training
                        Fold 0: finished training
                        Fold 0: Optimal threshold found (0.487)
                        Fold 0: Scoring trained pipeline
                        Fold 0: F1 score: 0.997
                Training and scoring on fold 1
                        Fold 1: starting training
                        Fold 1: finished training
                        Fold 1: Optimal threshold found (0.441)
                        Fold 1: Scoring trained pipeline
                        Fold 1: F1 score: 0.997
                Training and scoring on fold 2
```

```
                                Fold 2: starting training
                                Fold 2: finished training
                                Fold 2: Optimal threshold found (0.467)
                                Fold 2: Scoring trained pipeline
                                Fold 2: F1 score: 0.999
                        Finished cross validation - mean F1: 0.998
                        Finished cross validation - mean F1: 0.998
                CatBoost Classifier w/ Imputer:
                CatBoost Classifier w/ Imputer:
                        Starting cross validation
                        Starting cross validation
                                Training and scoring on fold 0
                                        Fold 0: starting training
                                        Fold 0: finished training
                                        Fold 0: Optimal threshold found (0.835)
                                        Fold 0: Scoring trained pipeline
                                        Fold 0: F1 score: 0.983
                                Training and scoring on fold 1
                                        Fold 1: starting training
                                        Fold 1: finished training
                                        Fold 1: Optimal threshold found (0.841)
                                        Fold 1: Scoring trained pipeline
                                        Fold 1: F1 score: 0.984
                                Training and scoring on fold 2
                                        Fold 2: starting training
                                        Fold 2: finished training
                                        Fold 2: Optimal threshold found (0.816)
```

```
automl.rankings
```

	id	pipeline_name	mean_cv_score	standard_deviation_cv_score	validation_score	per
0	4	LightGBM Classifier w/ Imputer + One Hot Encoder	0.998881	0.001109	0.999521	
1	5	Logistic Regression Classifier w/ Imputer + On...	0.998400	0.001210	0.999521	
2	6	XGBoost Classifier w/ Imputer + One Hot Encoder	0.998238	0.002223	0.999521	
3	7	Extra Trees Classifier w/	0.997599	0.001272	0.997121	

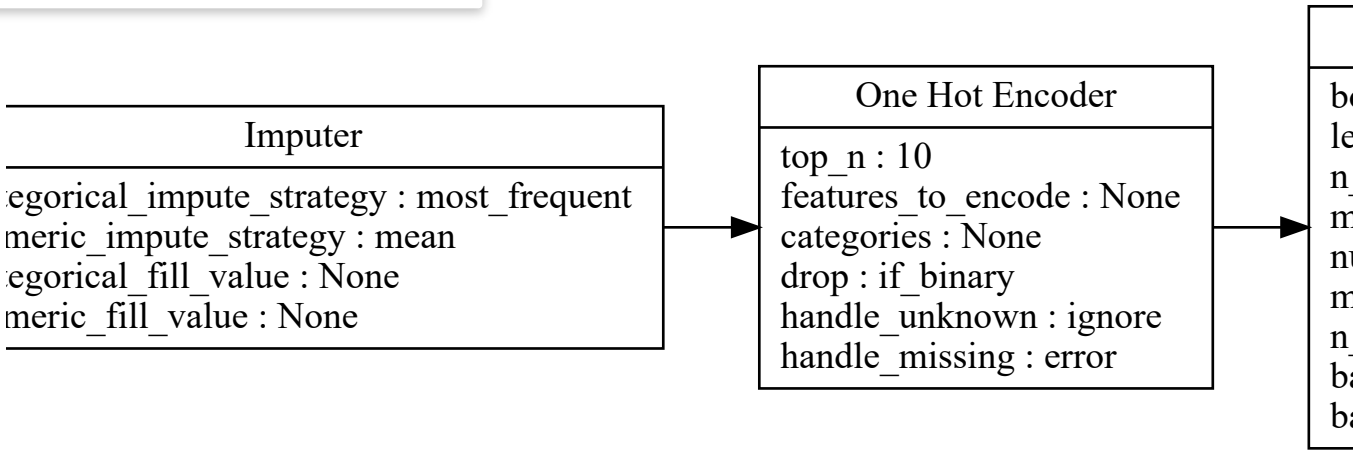
```
automl.describe_pipeline(4)
```

```
**************************************************

**************************************************
* LightGBM Classifier w/ Imputer + One Hot Encoder *
* LightGBM Classifier w/ Imputer + One Hot Encoder *
**************************************************
**************************************************


Problem Type: binary
Model Family: LightGBM


Pipeline Steps
Pipeline Steps
==============
==============
1. Imputer
1. Imputer
        * categorical_impute_strategy : most_frequent
        * categorical_impute_strategy : most_frequent
        * numeric_impute_strategy : mean
        * numeric_impute_strategy : mean
        * categorical_fill_value : None
        * categorical_fill_value : None
        * numeric_fill_value : None
        * numeric_fill_value : None
2. One Hot Encoder
2. One Hot Encoder
        * top_n : 10
        * top_n : 10
        * features_to_encode : None
```

```
            * features_to_encode : None
            * categories : None
            * categories : None
            * drop : if_binary
            * drop : if_binary
            * handle_unknown : ignore
            * handle_unknown : ignore
            * handle_missing : error
            * handle_missing : error
    3. LightGBM Classifier
    3. LightGBM Classifier
            * boosting_type : gbdt
            * boosting_type : gbdt
            * learning_rate : 0.1
            * learning_rate : 0.1
            * n_estimators : 100
            * n_estimators : 100
            * max_depth : 0
            * max_depth : 0
            * num_leaves : 31
            * num_leaves : 31
            * min_child_samples : 20
```

```
pipeline = automl.get_pipeline(4)
print(pipeline.parameters)
```

```
    {'Imputer': {'categorical_impute_strategy': 'most_frequent', 'numeric_impute_strategy':
```

Conclusion

We have a simple overview of some techniques and algorithms in machine learning. Furthermore, there are more and more techniques apply machine learning as a solution. The best model seem to

be the LightGBM Classifier w/ Imputer + One Hot Encoder model, which is the model with the mushrooms dataset The model generated a 'F1 score' of 0.998881.LightGBM Classifier w/ Imputer + One Hot Encoder is a great algorithm for classification problems to produce a predictive model.its default hyperparameters already return great results. So this project gives different ways of adressing the task with unbalanced data.Using Supervised learning Elastic Net Classifier, XGBoost Classifier , CatBoost Classifier , DecisionTreeClassifier,Logistic Regression and Random Forest.