**Mushroom Classification Using Different Classifiers**

****

| “Nature alone is antique and the oldest art a mushroom.” – Thomas Carlyle

Mushroom ! Creamy Mushroom Bruschetta, Mushroom Risotto, Mushroom Pizza, Mushrooms in a burger, and what not! Just by hearing the names of these dishes, people be drooling! Their flavour is one reason that takes the dish to the next level!

But have you ever wondered if the mushroom you eat is healthy for you ? From over 50,000 species of mushrooms only in North America, how will you classify the mushroom as edible or poisonous? Poisonous mushrooms can be hard to identify in the wild!

Let’s build different machine-learning models to classify the mushrooms into edible and poisonous !

**Introduction**

In this project, we will examine the data and build different machine learning models that will detect if the mushroom is edible or poisonous by its specifications like cap shape, cap colour, gill colour, etc. using difference classifier.

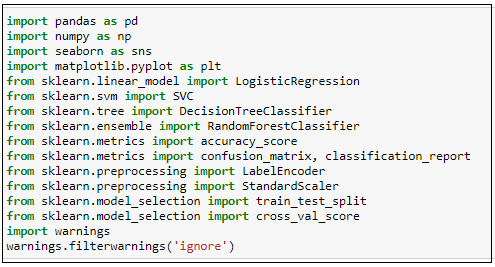
**Dataset**

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

**Libraries will be used**

* Numpy
* Pandas
* Seaborn
* Matplotlib
* Scikit-learn

**Importing the python libraries and Packages**

****

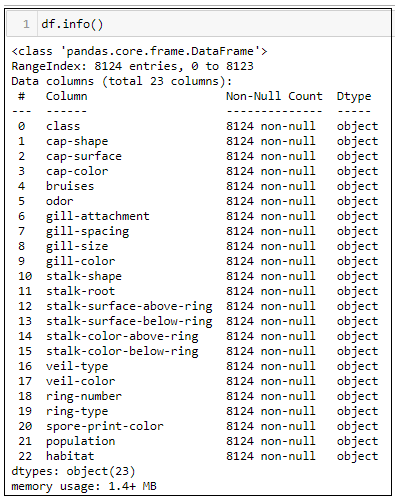
**Reading the csv file of the dataset**

Pandas read\_csv() function imports a csv file to DataFrame format.



**Examining the Data**

After importing the data, to learn more about the dataset, we’ll use.head() .info() and .describe() methods.



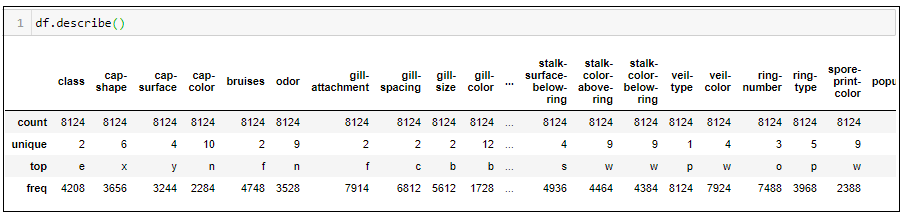
The .info() method will give you a concise summary of the Data Frame. This method will print the information about the Data Frame including the index dtype and columns types, non-values and memory usage.

**Descriptive Statistics**

The .**describe**() method will give you the **statistics** of the columns.

* Count shows the number of responses.
* Unique shows the number of unique categorical values.
* Top shows the highest-occurring categorical value.
* Freq shows the frequency/count of the highest-occurring categorical value.

Here is the output:



Unique occurrence of the class columns

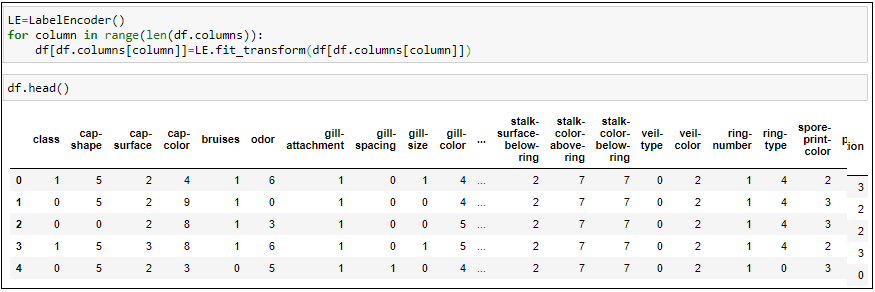


As we can see there are two unique values in the class columns of the dataset namely:

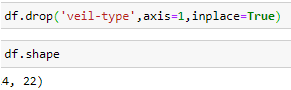
‘p’ -> poinsonous ‘e’ -> edible

**Pre-processing**

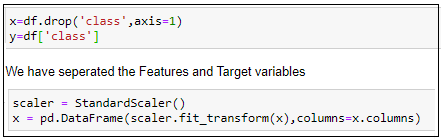
Let’s convert the dataset to make sure all the object types are converted into numerical values.



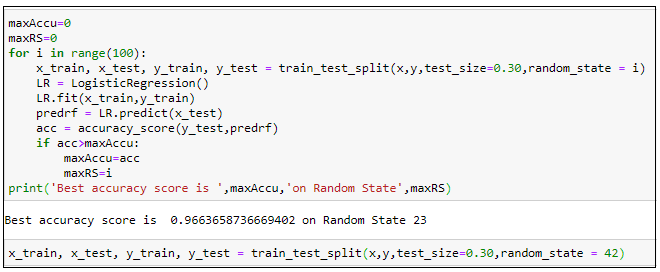
Dropping the column ‘veil-type’ which doesn’t have any values



**Splitting the dataset to build the machine learning model**

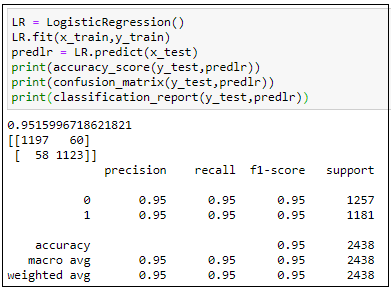
****

**Finding the best random states**

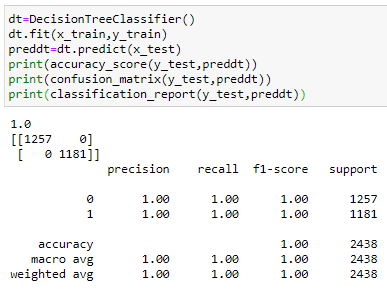
****

We found that the best random status to build the model is 42

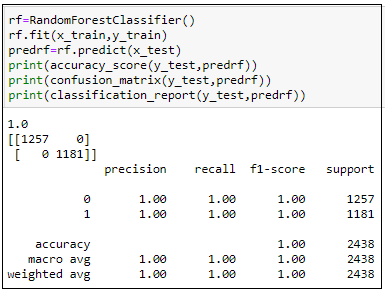
1. LogisticRegression



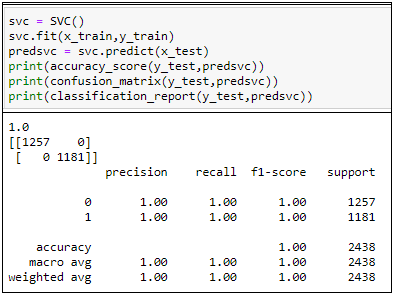
1. DecisionTreeClassifier



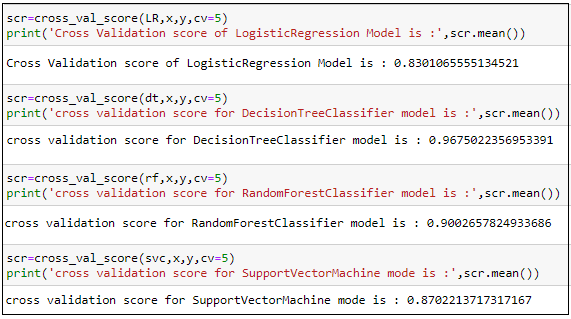
1. RandomForestClassifier



1. SupportVectorMachine



Luckily we have received 100% accuracy in all the models except Logistic Regression which is 95%. However, let’s try with the cross validation to check whether there is overfitting or under fitting.



Finally we came to a conclusion that the DecisionTreeClassifier is the best fitted model which has the least difference between the actual accuracy score and cross validation score.

**Conclusion**

From the model development, we saw that our train and test data is balanced.

Most of the classification method hits 100% accuracy with this dataset.

Woohoo! Congratulations!!! We can now eat healthy mushrooms!! Yay!!