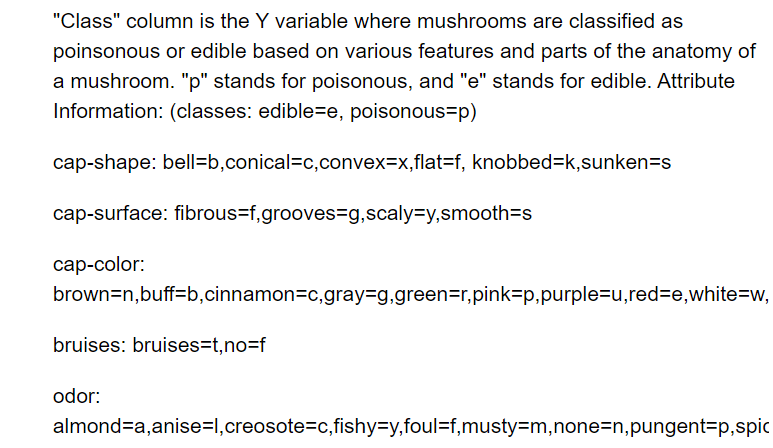
Website content:

**Special considerations noted in the dataset**:

The dataset is comprised of entirely categorical classifications for various attributes of mushroom species. Each species is identified as “definitely edible, definitely poisonous, or of unknown edibility and not recommended (“Mushroom Classification.” Kaggle.com). All attributes are assigned unique letter classifications that determine the different features. For example, on “Cap Shape,” the letter “b” stands for a bell shaped cap top of the mushroom.

**Shape of the dataset utilized**:

The dataset is clearly a supervised learning set with a clear “Class” column that determines whether a mushroom is edible or poisonous. There are 22 features for mushroom attributes.



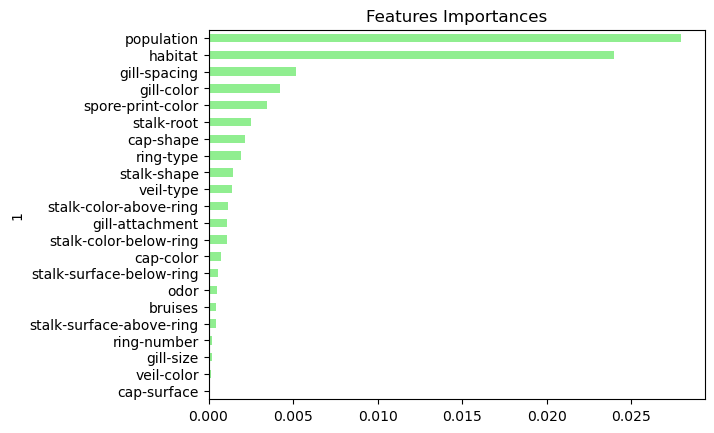


**Data Analysis Process**:

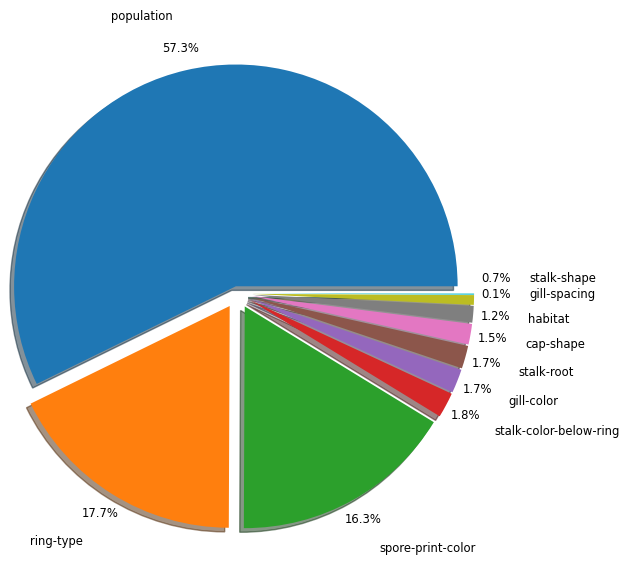
OnehotEncoder for the X Matrix & LabelEncoder for the Y Label. Both are needed to maintain attribute integrity and correctly convert the categorical values to numerical inputs.

Model chosen was the Rand Forest due to its strength and robust nature regarding various categorical classifications that would lead to a “yes” or “no” decision, or a “True” or “False” decision making process, ie: If a mushroom had a certain attribute, yes or no.

First Model utilized all 22 features, and then an analysis was conducted via confusion matrix, accuracy score testing, and feature importance ranking.



The optimized Model utilizes the top 10 features while maintaining accuracy score of 1 and for simplicity in the machine model that powers our application.



The Top 10 Features show an overwhelming significance in population, spore-print-color, and ring-type as the highest indicators of a mushroom's deadliness. These top features are optimized in determining the edibility of mushrooms in the model.