

## Mushrooms – Edible, poisonous, or unknown?

There exists a mantra in the world of mycologists: “Never eat a mushroom unless you can positively identify it.” With approximately 14,000 species of mushrooms that we know of, the process of classifying them seems so daunting that someone’s better off sticking to store bought mushrooms for their culinary needs. Unless, a supervised machine learning algorithm could be utilized to assist in classification.

The **Mushroom Data Set**, complimentary of the University of California, Irvine Machine Learning Repository, was chosen. There are 8,124 instances with 2,480 missing the same attribute and for consistency would be removed from the data set. All together, there are twenty two attributes:

- Cap shape = {bell, conical, convex, flat, knobbed, sunken},
- Cap surface = {fibrous, grooves, scaly, smooth},
- Cap color = {brown, buff, cinnamon, gray, green, pink, purple, red, white, yellow},
- Bruises = {T, F},
- Odor = {almond, anise, creosote, fishy, foul, musty, none, pungent, spicy},
- Gill attachment = {attached, descending, free, notched},
- Gill spacing = {close, crowded, distant},
- Gill size = {broad, narrow},
- Gill color = {black, brown, buff, chocolate, gray, green, orange, pink, purple, red, white, yellow},
- Stalk shape = {enlarging, tapering},
- Stalk root = {bulbous, club, cup, equal, rhizomorphs, rooted},
- Stalk surface above ring = {fibrous, scaly, silky, smooth},
- Stalk surface below ring = {fibrous, scaly, silky, smooth},
- Stalk color above ring = {brown, buff, cinnamon, gray, orange, pink, red, white, yellow},
- Stalk color below ring = {brown, buff, cinnamon, gray, orange, pink, red, white, yellow},
- Veil type = {partial, universal},
- Veil color = {brown, orange, white, yellow},
- Ring number = {none, one, two},
- Ring type = {cobwebby, evanescent, flaring, large, none, pendant, sheathing, zone},
- Spore print color = {black, brown, buff, chocolate, green, orange, purple, white, yellow},
- Population = {abundant, clustered, numerous, scattered, several, solitary},
- Habitat = {grasses, leaves, meadows, paths, urban, waste, woods}.

Finally, there is a binary classification value of “edible” or “poisonous”.

For this problem, I propose harnessing two different supervised learning methods: Decision Tree Classification and Naïve Bayes Classification. Both of these methods are ideal candidates for solving classification problems. The main point of interest will be analyzing the performance of these two algorithms and determining the more successful approach. Google’s Colaboratory will be used to conduct the experiment. Colaboratory was selected because its hassle-free deployment and the universal nature of the software simplifies submitting deliverables. Furthermore, the Scikit-Learn library will be utilized for its easy-to-use and well-documented code base.