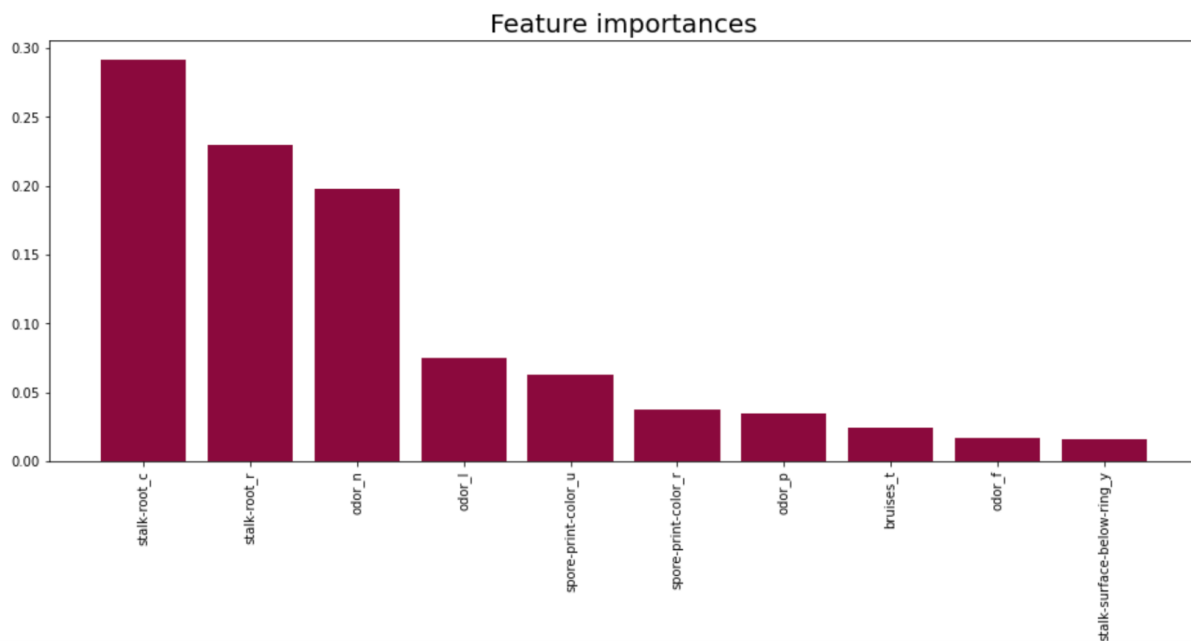


Juliana McCausland
Metis ML Classification Module
MVP

Goal: classify poisonous and edible mushrooms, and identify features of these classes

My baseline, which only uses a few of the nearly-100 features in my dataset, has a logistic regression test score of 98.40% and a KNN test score of 99.02%. When running these same models on the entire dataset, I get 100% scores. I have also identified some of the most important features after fitting to an XGBClassifier model, as seen below:



From left to right:

Stalk-root_c = club-shaped stalk root of the mushroom

Stalk-root_r = rooted stalk root

Odor_n = odorless

Odor_l = anise odor (licorice-like)

Spore-print-color_u = purple spore print

Spore-print-color_r = green spore print

Odor_p = pungent odor

Bruises_t = bruised (the alternative is 'no bruises')

Odor_f = foul odor

Stalk-surface-below-ring_y = scaly surface below ring

We now have an idea of which features are important when identifying poisonous mushrooms. It seems like stalk root and odor are highly important. The next steps will be to remove the most important features and evaluate the accuracy of those models, and look at the other features

that play a role in identifying poisonous vs edible mushrooms. It is especially important to understand specific features of poisonous mushrooms for individuals (survivalists?) who are out in the real world, perhaps scavenging for food, and only have a limited number of 'features' that they are able to identify using their own sight and sense of smell.