

Edibility Classification of Mushrooms

By: Ben Sturm

Minimum Viable Product Summary

Domain

The project I have chosen for Project McNulty is associated with the classification of mushrooms based on their edibility, that is being poisonous or non-poisonous. This data is available on the [UCI repository](#). I think this will be an interesting topic because I'm an amateur mushroom hunter myself and perhaps can learn something from this dataset. I also very much like the fact that the data is already given to us in a nice, easy to read format, such that I'll be able to start building my models quickly.

Data

This dataset contains 8124 instances of different mushrooms all within two mushroom families, Agaricus and Lepiota. There are 22 total attributes for each of these mushrooms all of which being categorical, which are given in the list below:

- Cap-shape, Cap-surface, Cap-color
- Bruises?
- Odor
- Gill-attachment, Gill-spacing, Gill-size, Gill-color
- Stalk-shape, Stalk-root, Stalk-surface-above-ring, Stalk-surface-below-ring
- Stalk-color-above-ring, Stalk-color-below-ring
- Veil-type, Veil-color
- Ring-number, Ring-type
- Spore-print-color
- Population
- Habitat

The great thing about this data set is that it is very balanced, in that there are relatively equal number of poisonous mushrooms and non-poisonous ones. I plan to use all features for this data set.

Known unknowns

At this point, I don't know how well a classification algorithm will perform, but based on my limited knowledge of mushroom identification, the attributes in this data set are used to help identify mushrooms. Given that, I think the data should contain the information necessary to

classify the mushrooms and thus pick out those that are edible and poisonous. I believe I will be able to start with a relatively simple LogisticRegression model to begin. Then I plan to explore the performance of other models such as Naïve Bayes, Decision Trees, Gradient Boosted Decision Trees and so on.