# Project 8: Applied Theory & Practice I, Mushrooms

## Preprocessing the Data

## Can you generate summary statistics that help describe the data?

The data for the mushrooms is pretty well discretized; there are no continuous values, and very few ordinal values. So the most interesting summary statistics for this data are the quick statistics on amounts and ratios of different classes. There are 8124 data points in the data: 51.8 are poisonous, and 48.2 are edible. So, there is quite a bit of data, and an even distribution. Continuing to explore the data in WEKA’s explorer there are some other interesting things we should be aware of. There are 31% missing attributes in the stalk-root column. All mushrooms have the same veil type. We can take that out of our considerations as the data is mined later.

One the odor attribute is very interesting because it very accurately separates the data into poisonous and edible. This could be useful for classification later.

## Can the edible and poisonous data objects be distilled into groups?

## Can a classification model be created that can predict whether a mushroom is edible or poisonous?

Yes, one-way to do this is a rule-based classifier.

## Do any anomalies exist in the dataset?

## Can any association rules be generated from this dataset?