# Model Card – Mushroom Edibility

Model Details

Decision Tree, Created by Dr V. Crockett, 28/10/2022

The model was created from a data set containing 22 features about 8124 mushrooms to classify whether they are potentially edible or not.

Intended Use

Due to the unrepresentative nature of the data, missing features, and high stakes (poisoning) this model is not to be used to determine mushroom edibility in any practical way.

This model is intended as a tutorial for Machine Learning methodology and to demonstrate the creation of a model card.

Factors

Variables that were used following construction of the tree: cap surface, odor, spore print color, stalk color below the ring, stalk root shape.

The model was restricted by a minimum of 10 observations to complete a split (and therefore 4 to form a final decision).

Splits were restricted by setting the complexity parameter to 0.0001, equivalent to needing to improve R2 by at least 0.0001. This pre-prunes for computational efficiency.

Training and Evaluation Data

6436 (80%) of the observations were used for training the model.

1688 (20%) of the observations were set aside for testing.

Data source: [UCI Machine Learning Repository: Mushroom Data Set](https://archive.ics.uci.edu/ml/datasets/Mushroom)

Diagram

Description automatically generated

Metrics

Each to 4 significant figures: Accuracy: 99.94% Precision: 1 Recall: 0.9988

False Negative Rate: 0.001172 False Positive Rate: 0

Caveats and Recommendations

The data source is based on 23 species of gilled Agaricus and Lepiota North American mushrooms – it is less likely to be accurate outside of this geographical area and for a wider variety of mushrooms. Additionally, data was gathered in 1981 and it is possible that environmental and genomic changes since then may mean that this data is less representative of mushrooms today. Further research and data gathering is recommended to improve the quality of the data set if the model usage is extended.

An improved data set would also contain at least the following key measurements that are presently not included: surrounding habitat, time of year, growth pattern, availability and level of local labelling expertise. Additional features should be discussed with a Fungi domain expert prior to future data gathering.