Classification of Mushroom Edibility

Team Conifer: Dylan Cordaro, Schinella D’Souza, Teresa Yu

Motivation

- Mushroom hunting/foraging is a popular practice throughout the world
  - Mushrooms play an important role in many cultures and cuisines
- Important to distinguish between safe, edible mushrooms from poisonous, inedible ones
  - Poisonous mushrooms have significant adverse health risks and can be lethal
- Challenges
  - Many edible mushrooms have poisonous lookalikes
  - No obvious “rule” to distinguish between the two classes

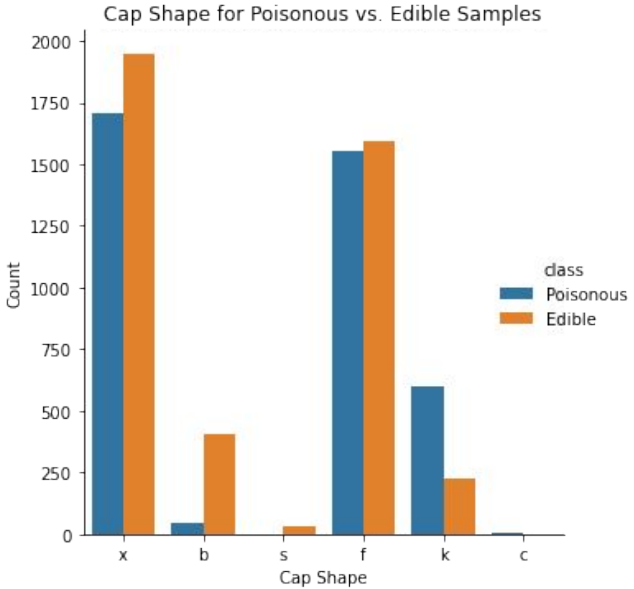
Real and false chanterelles
Image credit: Jamie Kunka via Youtube

Data Collection

- Dataset obtained from the UCI Machine Learning repository via Kaggle
- Descriptions of 8124 hypothetical samples from 23 species of gilled mushrooms
  - Samples drawn from *The Audubon Society Field Guide to North American Mushrooms*
- Attributes described for each sample:
  - Edibility (edible vs. poisonous)
  - Cap shape, surface, color
  - Bruising
  - Odor
  - Gill attachment, spacing, size, color
  - Stalk shape, root, surface texture, color
  - Veil type, color
  - Ring number, type
  - Spore print color
  - Density
  - Habitat

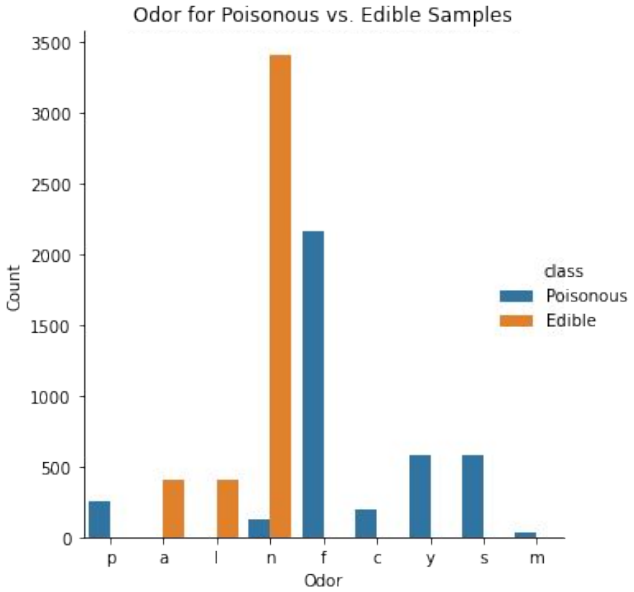
Exploratory Data Analysis

- 8124 samples, 22 features
- Plotted histograms for various features, given each class



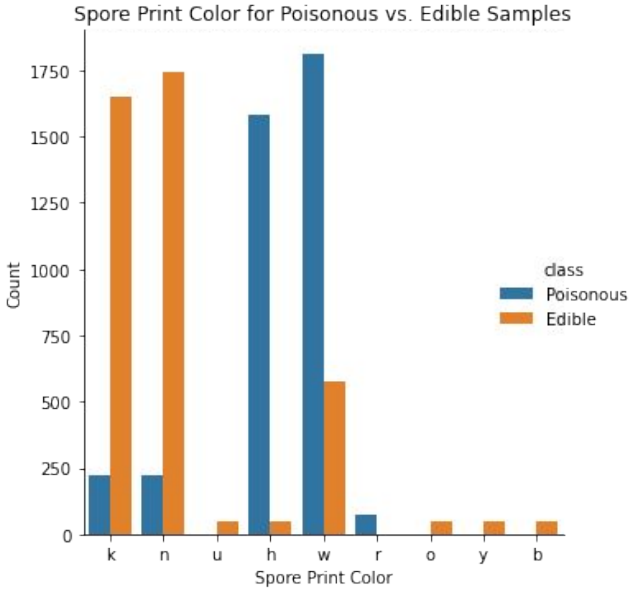
Exploratory Data Analysis

- 8124 samples, 22 features
- Plotted histograms for various features, given each class



Exploratory Data Analysis

- 8124 samples, 22 features
- Plotted histograms for various features, given each class



Exploratory Data Analysis

- 8124 samples, 22 features
- Plotted histograms for various features, given each class

Conclusion

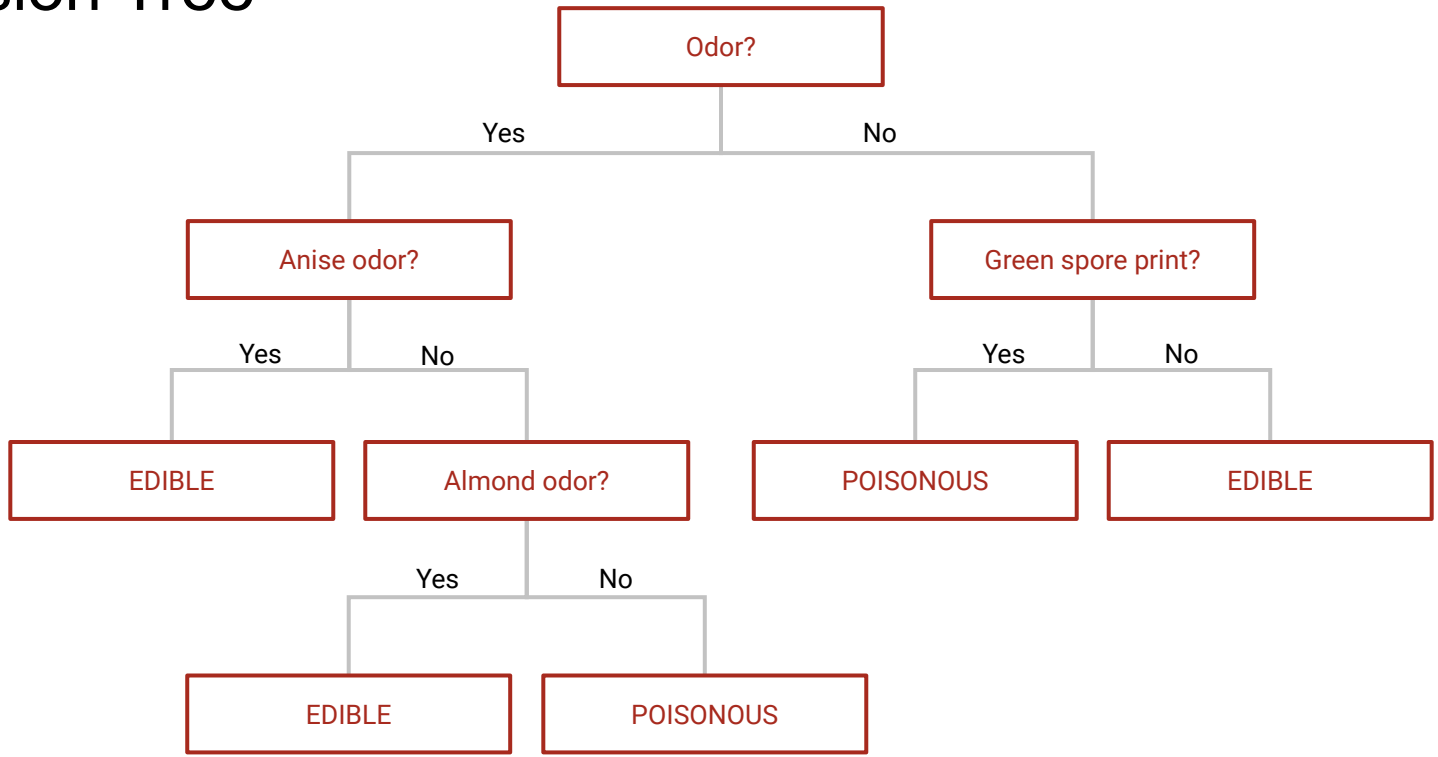
- Odor, spore print color features appear to best distinguish edibility

Modeling Approach

- Categorical data
- Models that we tried:
  - Decision tree
  - Naïve bayes classifier
  - Logistic regression
- Selected odor and spore print color features to fit models
- Used 75-25 train/test split and 5-fold cross-validation
- Evaluated performance using test accuracy

Conclusion: decision tree of depth 3 is an effective model

Decision Tree



Results and Recommendations

- Model conclusions:
  - Anise or almond odor means EDIBLE
  - Any other odor means POISONOUS
  - Odorless and green spore print means POISONOUS
  - Odorless and non-green spore print means EDIBLE
- Our model had ~99% test accuracy
  - Errors: misclassifying odorless and white spore print mushrooms as being edible, when they're actually poisonous
- Other conclusions:
  - Need to be careful about odorless and white spore print samples
  - Still best practice to consult field guides and experts if ever unsure of exact species of a mushroom sample

False chanterelles are poisonous, odorless, and have white spore prints. Image credit: David Nicholls via NatureSpot