

Problem Identification

What properties of a mushroom guarantee its edibility?
toxicity?

- ① Context: Mushroom hunting is a potentially very lucrative, but eating the wrong mushroom can be very poisonous.
- ② Criteria for success:
 - low false positive rate (would like to keep most edible ones)
 - very low false negative rate (going to the ER would be bad)
- ③ Scope of solution space: Focus on mushrooms in North America. Look at categorical data determined by the phenotype of a mushroom, e.g. its cap, gills, etc.

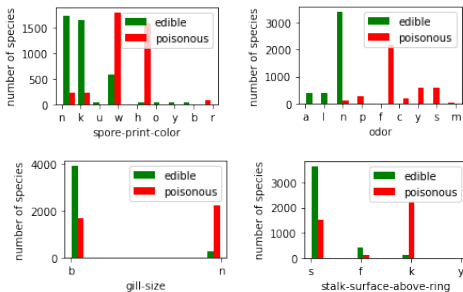
Problem Identification

- ④ Constraints within solution space:
 - The data is 30 years old (may not be up to date)
 - Model may not be accurate on other continents
- ⑤ Stakeholders to provide key insight:
 - mushroom foragers
 - mycologists
- ⑥ Key data source:

Kaggle dataset from a 30-year-old UCI Machine Learning repository – describes hypothetical samples corresponding to 23 species of gilled mushrooms from The Audubon Society Field Guide to North American Mushrooms (1981).

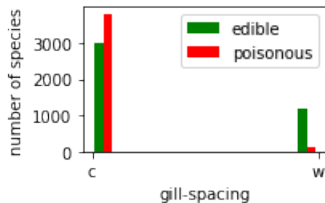
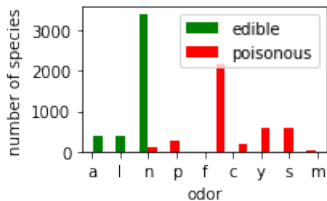
Recommendation and key findings

We are able to determine the four traits most strongly associated with poisonous mushrooms. They are, in order of importance, green spore prints (r), creosote odors (c), narrow gills (n), and silky stalk surfaces above the ring (k). This agrees with the histograms in our exploratory data analysis:



Recommendation and key findings

We are also able to determine the four traits most strongly associated with edible mushrooms. They are, in order of importance, almond odors (a), anise odors (l), no odors (n), and crowded gill spacing (w). This agrees with the histograms in our exploratory data analysis:



Modeling results and analysis

SVC coefficients for poisonousness:

spore-print-color_r	1.591870
odor_c	1.045295
gill-size_n	0.799876
stalk-surface-above-ring_k	0.778354
population_c	0.762460
⋮	⋮
stalk-root_r	−0.583111
gill-spacing_w	−0.621908
odor_n	−0.844349
odor_l	−1.022896
odor_a	−1.022902

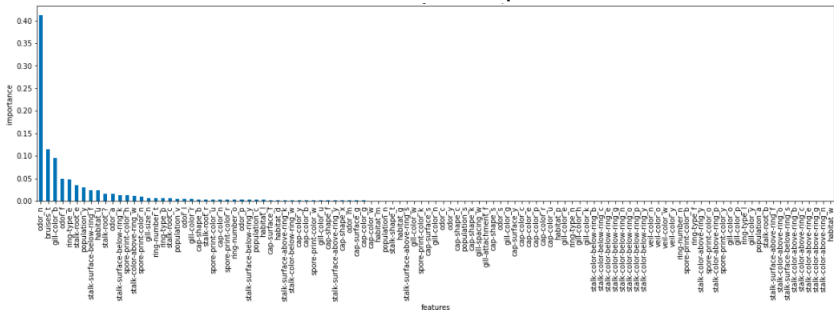
Modeling results and analysis

Logistic regression coefficients for poisonousness:

<code>gill-size_n</code>	3.125447
<code>stalk-surface-above-ring_k</code>	3.101673
<code>bruises_t</code>	2.537706
<code>stalk-root_c</code>	1.694761
<code>ring-type_p</code>	0.682597
<code>:</code>	<code>:</code>
<code>spore-print-color_k</code>	-3.908160
<code>spore-print-color_n</code>	-3.965740
<code>odor_l</code>	-5.063126
<code>odor_a</code>	-5.108882
<code>odor_n</code>	-6.226523

Modeling results and analysis

Random forest feature importances:



Summary and conclusion

- The most poisonous mushroom features are green spore prints, creosote odors, narrow gills, and silky stalk surfaces above the ring.
- The features most strongly associated with edible mushrooms are almond odors, anise odors, no odors, and crowded gill spacing.
- Other important features determining mushroom class are bruises, buff gill color, and club stalk roots.