Mushroom: To eat or not to eat?

RACHEL CHIANG

What are some qualities of Mushrooms?

Cap

Scales

Stem

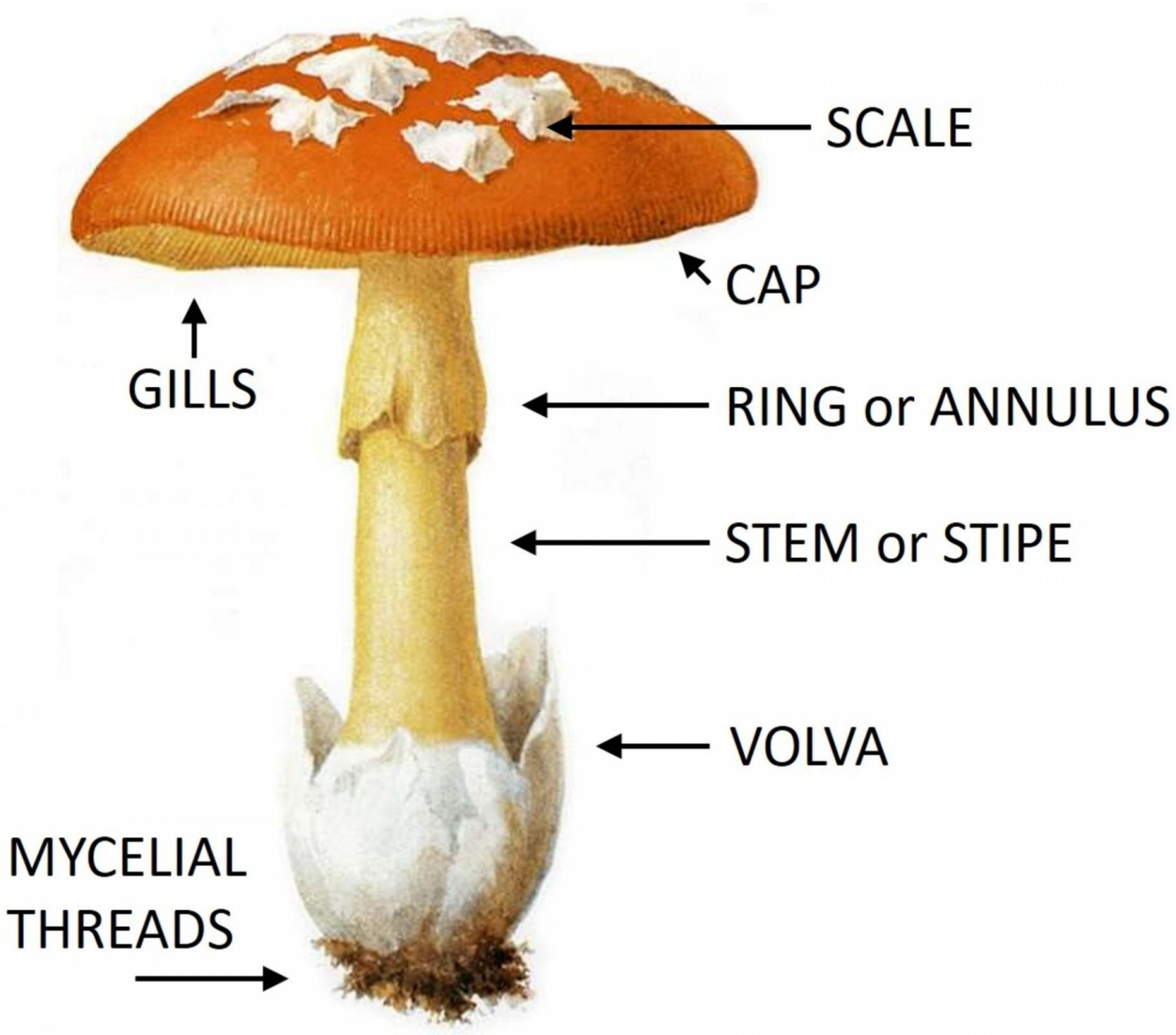
Gill

Ring

Habitat

Season

Bruising or bleeding



This Dataset

Numerical measurements

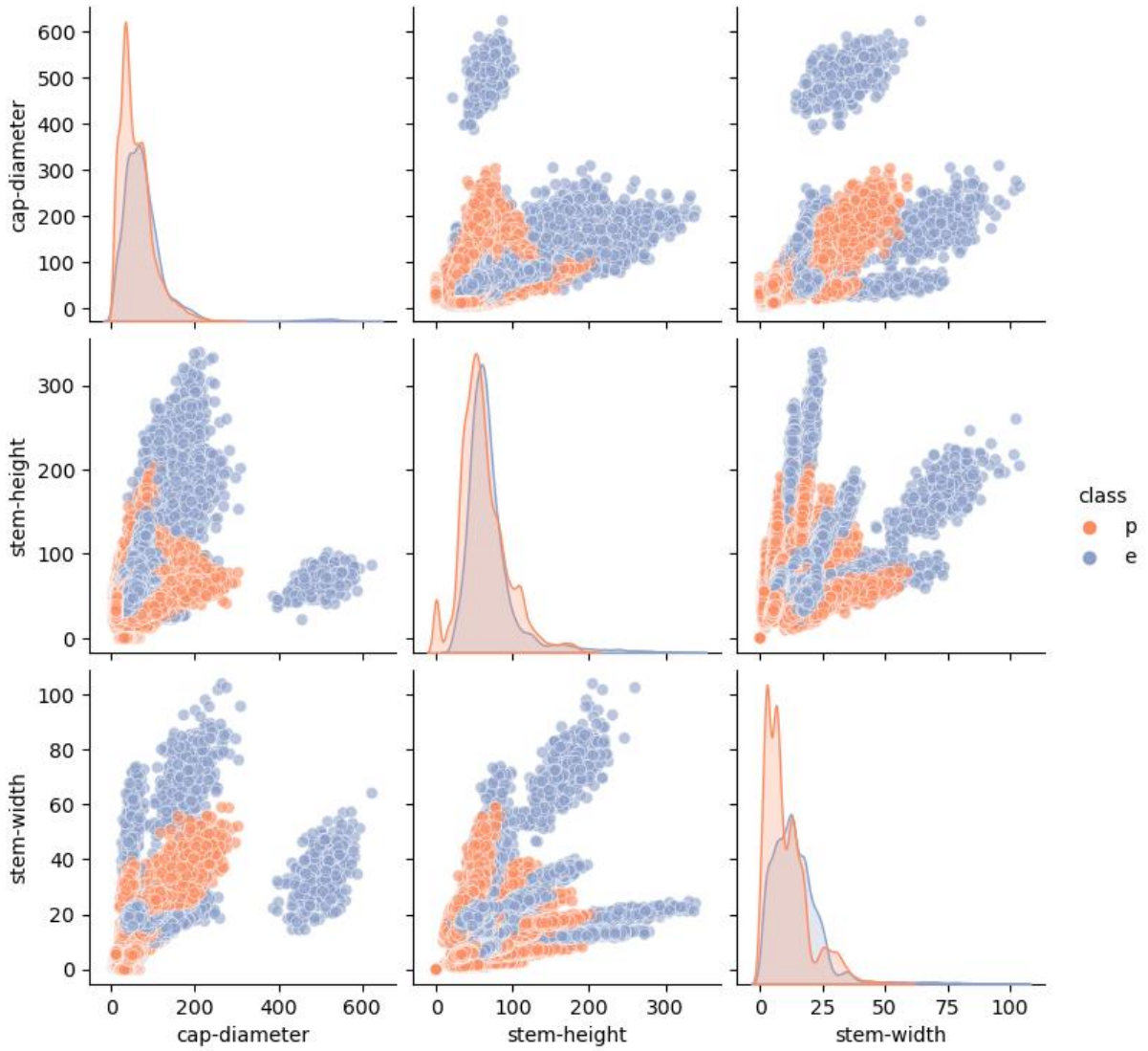
- Cap diameter
- Stem height and width

Descriptive qualities

- Cap shape, surface, and color
- Gill attachment, spacing, and color
- Stem color
- Ring type and existence

Behavioral

- Habitat
- Season
- Bruising or bleeding



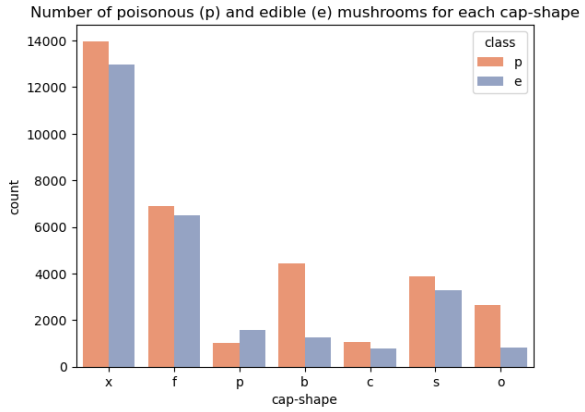
In the Numbers

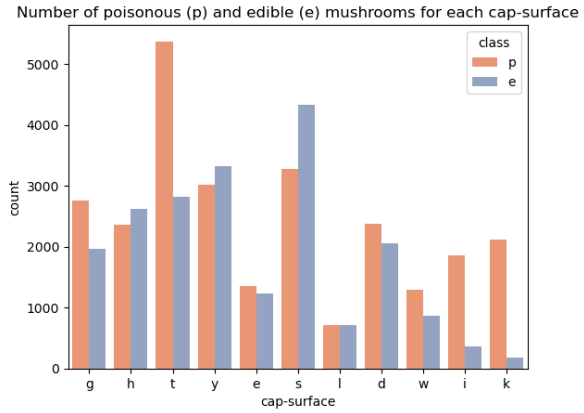
Smaller tend to be poisonous

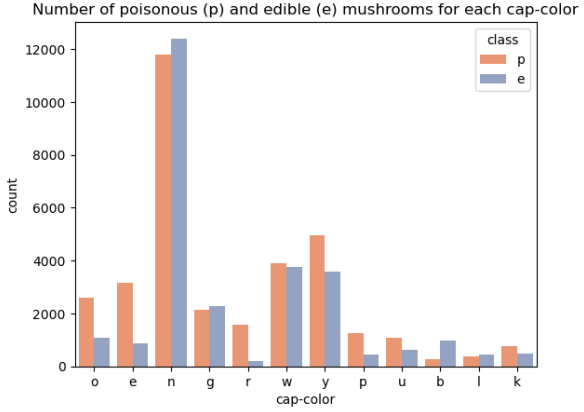
Larger tend to be edible

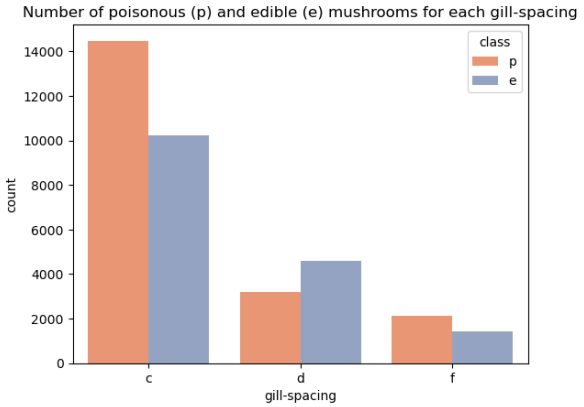
Caps

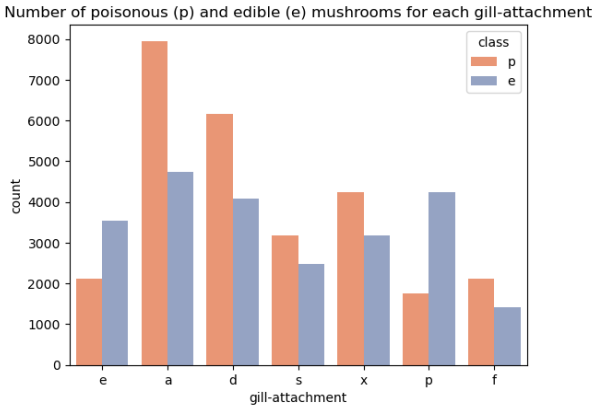
Poisonous	Edible
Bell or other cap shapes	Spherical cap shapes
Sticky, silky, fibrous cap surface	
Bright or vivid colors (green, pink, orange, red)	Buff cap color

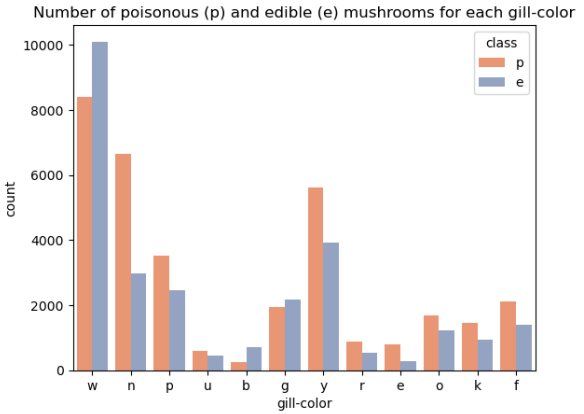








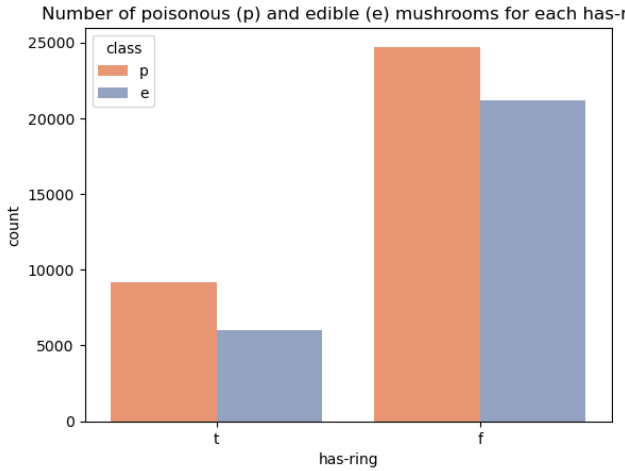


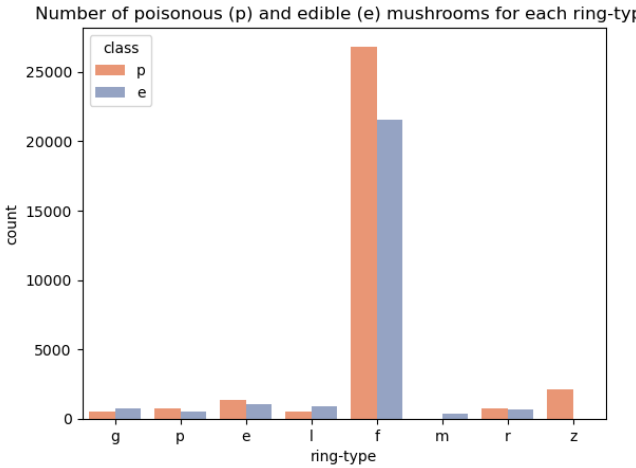


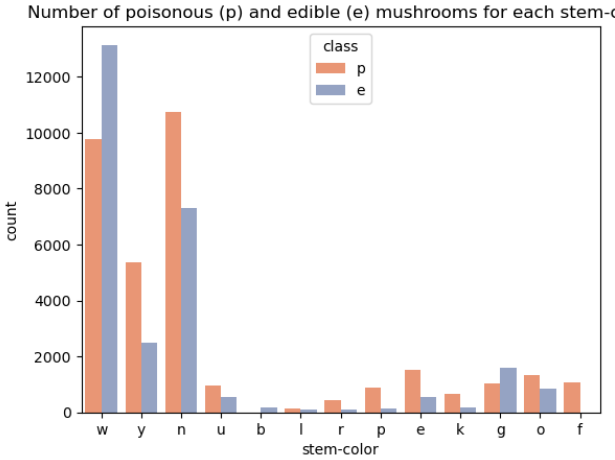
Poisonous	Edible
Closer-spaced gills	Distant-spaced gills
Decurrent or adnate gill attachments	Pores (not gills)
Bright or vivid colors (red, yellow, green, pink, …)	Buff gill color

Gills

Rings and Stems

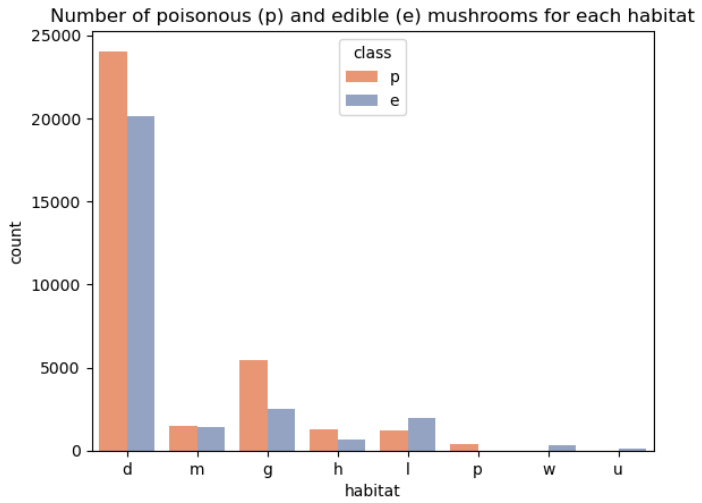


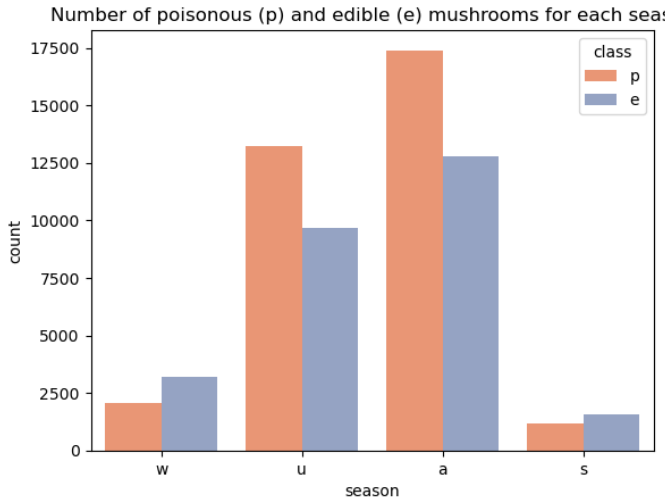


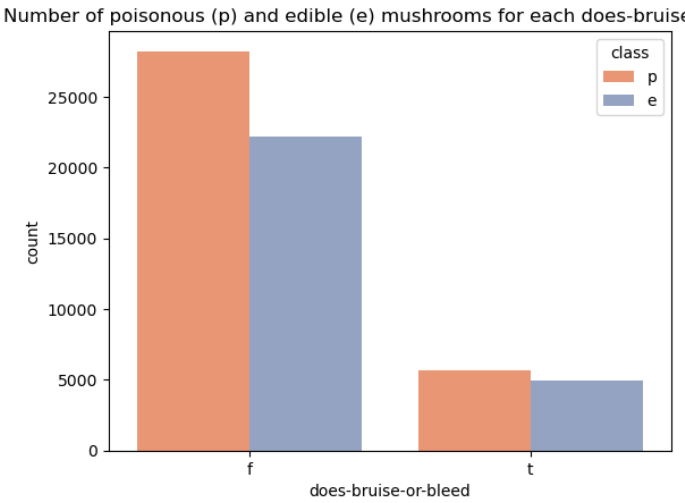


Poisonous	Edible
Simply having a ring or not does not seem to give much information.	
Zone type ring	Movable type ring
Bright or vivid colors (red, green, pink, …) and black	Buff stem color

Behaviors







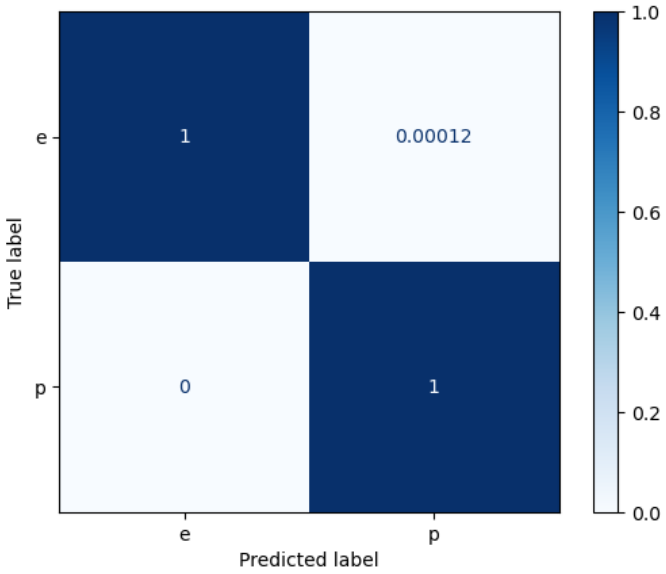
Poisonous	Edible
Path, heaths, and grass mushrooms	Urban and waste mushrooms
Summer and autumn	Winter
May be too difficult to tell from just bruising or bleeding	

Models

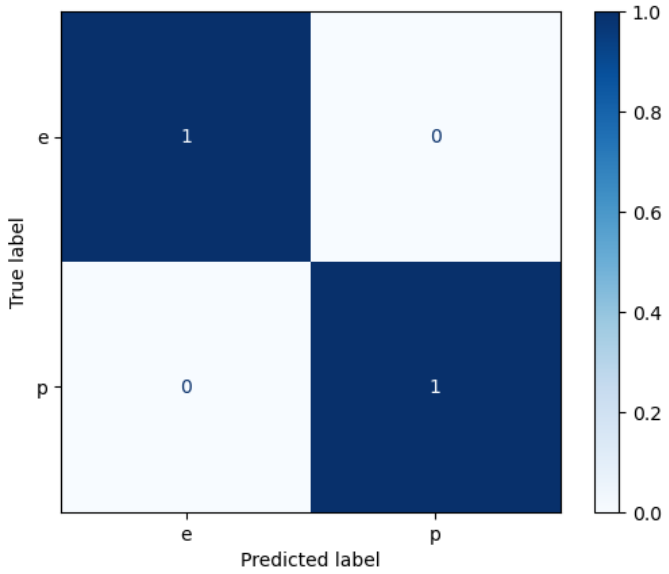
Model	Training Accuracy	Testing Accuracy
Logistic Regression	0.7826097127350987	0.7790513618252278 8
K Nearest Neighbors (k=1)	0.9999532163742689	0.9999454178265379
Random Forest (Gini, 20% max features, 60 estimators)	0.9999766081871344	1.0
Gradient Boosting (Hist, learning rate 0.7, 40 max iterations)	0.9999298190889944	0.9998908356530757

Confusion Matrices

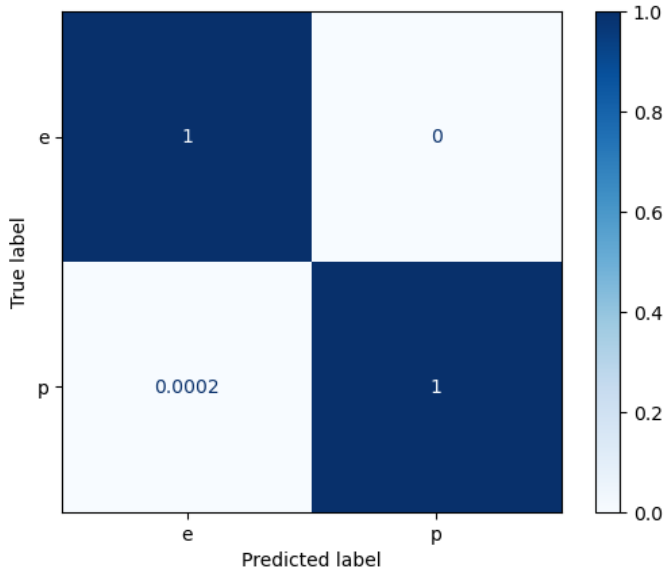
KNN



Random Forest



Gradient Boosting



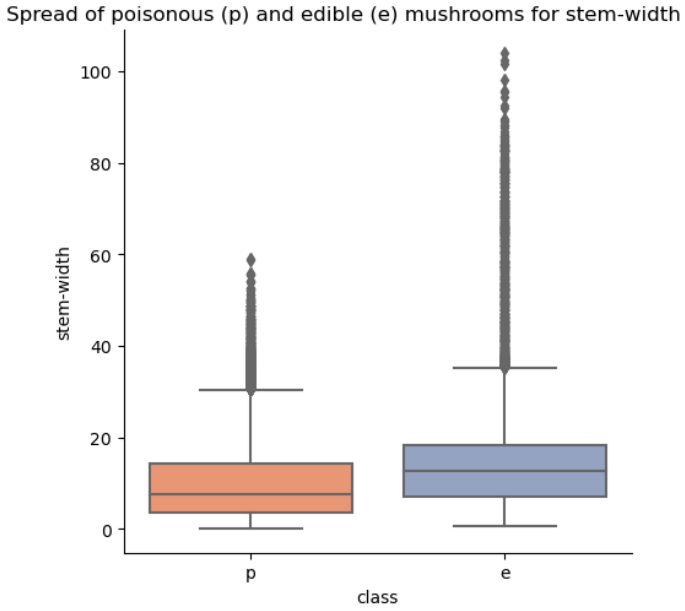
Speed

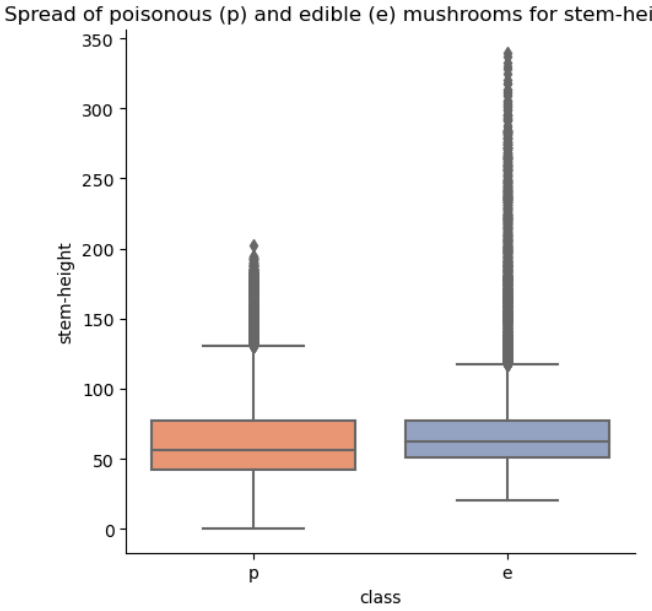
MODEL	COMPUTATION TIME [MS]
KNN	8437.627077102661
Random Forest	130.6772232055664
Histogram-based Gradient Boosting	21.938323974609375

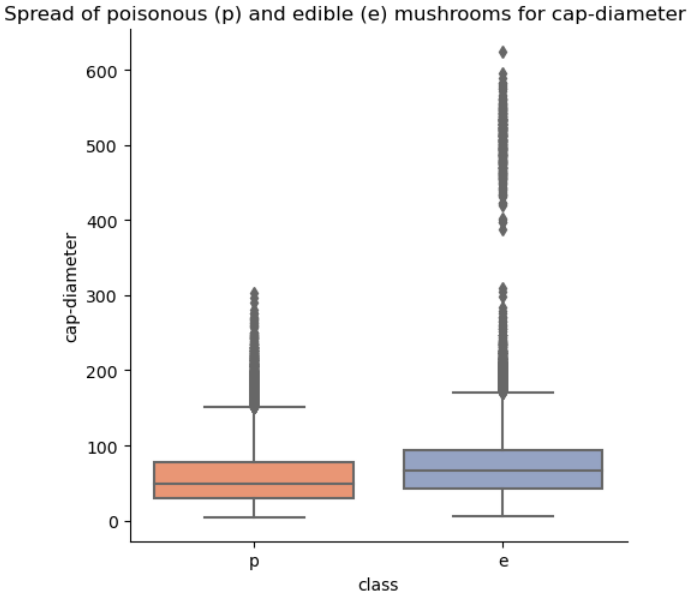
Interpretation

Feature	Importance
Stem width	0.097118
Stem height	0.071449
Cap diameter	0.056837
Stem color (white)	0.040072
Gill spacing (distant)	0.031649

The Three Numerical Features







A narrower question…

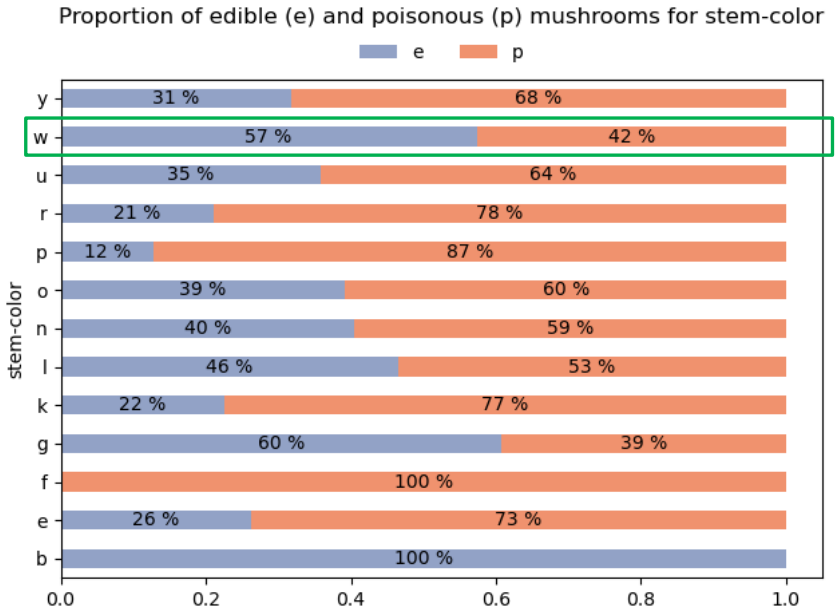
CAN WE FIND GOOD *VISUAL* FEATURES THAT HUMANS CAN USE TO DECIDE IF A MUSHROOM IS POISONOUS OR NOT?

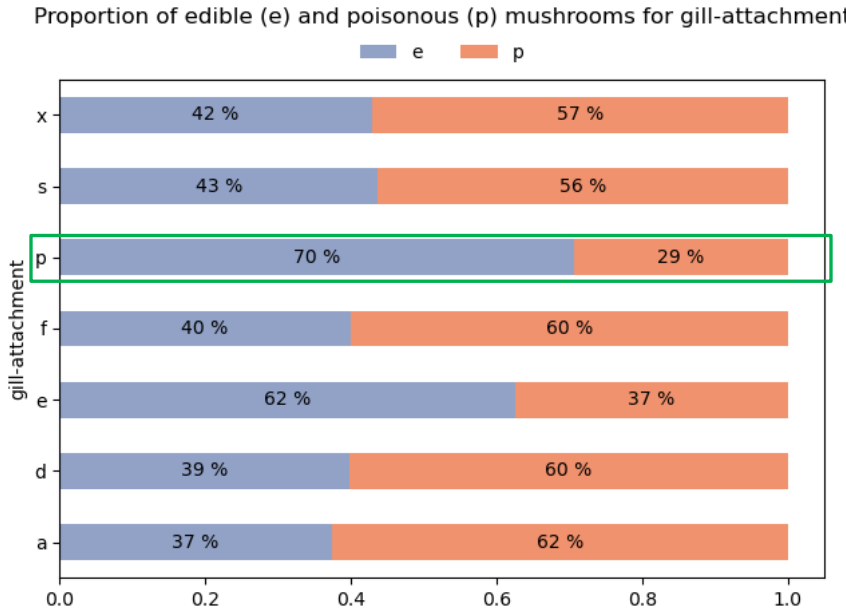
A Categorical RF Model

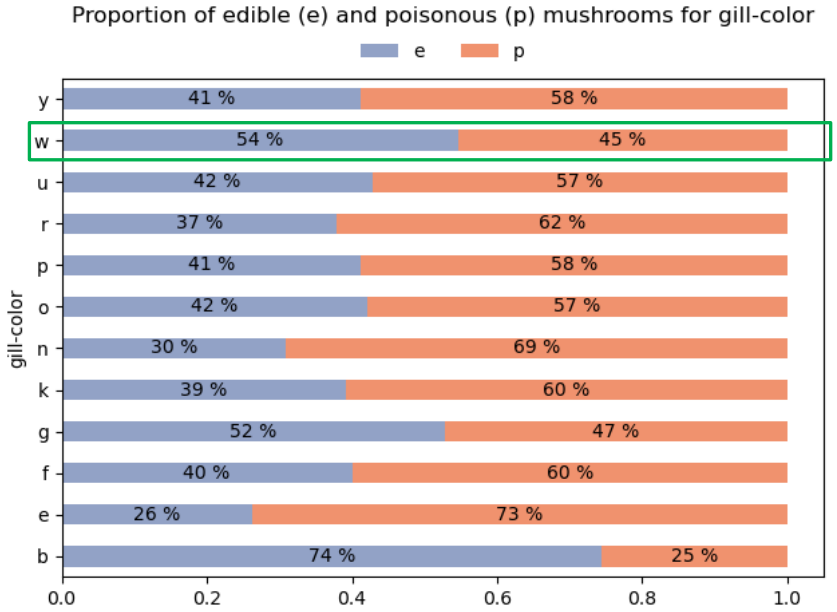
Training Accuracy	Testing Accuracy
0.9965378393150186	0.996288412204574

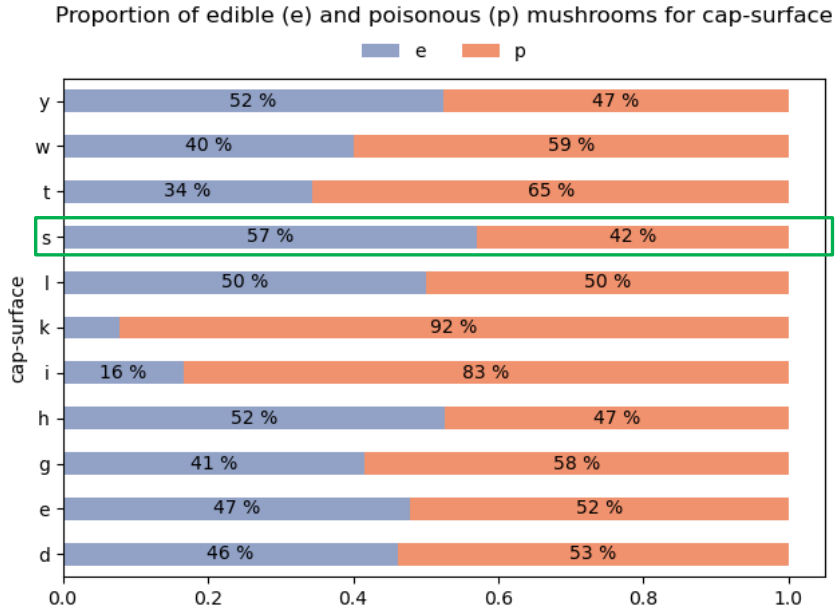
Feature	Importance
Stem color (white)	0.041744
Gill attachment (pores)	0.030955
Gill color (white)	0.030158
Cap surface (smooth)	0.029580
Gill spacing (distant)	0.028995

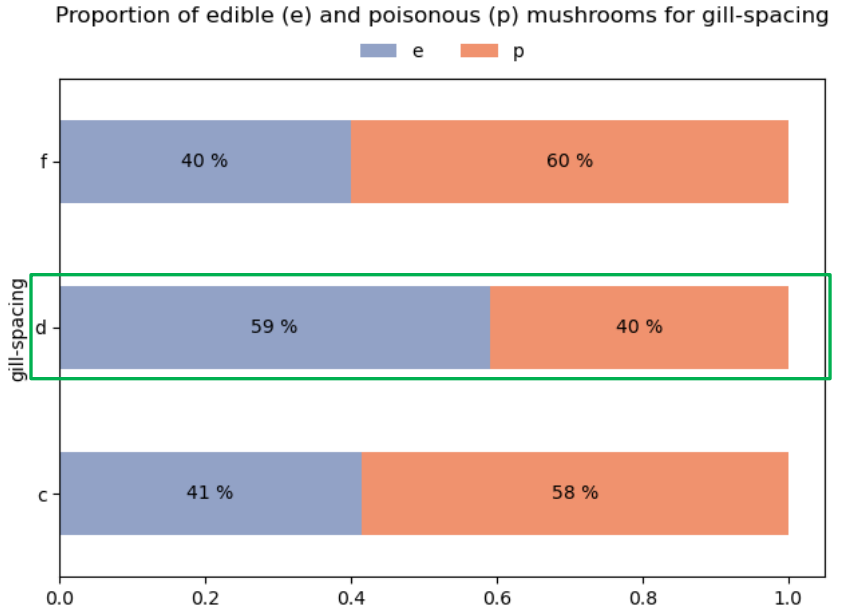
Stem Color, Gill Features, and Cap Surface











An answer?

It is probably edible if...

- the stem is white,
- it’s a porous mushroom,
- the gills are also white,
- the cap is smooth,
- and the gills are spaced apart.

Conclusion

Models

- Random forest
- Histogram-based gradient boosting

Looks that (don’t) kill

- Large, smooth, white (esp. stem and gill), porous mushrooms

Be safe

Sources

Mushroom dataset (UCI/Kaggle)

Mushroom parts diagram (image)

Giant puffball (image)