The Prediction of Mushroom Edibility Using Machine Learning

Benjamin W. Sturm
bwsturm@gmail.com
May 16, 2018

Hunting edible mushrooms in the wild can be an exciting and rewarding process

Hunter-Gatherers

The Trophy

Wouldn’t it be great if you could determine whether a mushroom you encounter in nature is poisonous or not with machine learning?

Agaricus Campestris (Non-poisonous)

Edibility

Edible, of good flavor, but lacking the solid texture of an *Agaricus bitorquis*.

Agaricus Californicus (Poisonous)

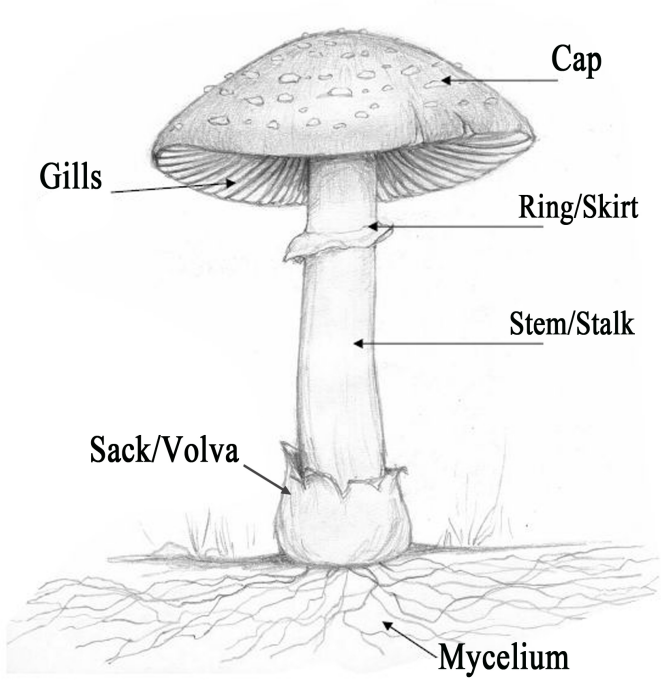
Edibility

Mildly toxic. Causes gastrointestinal upsets for most people although some individuals are able to eat it without negative effects.

Wouldn’t it be great if you could determine whether a mushroom you encounter in nature is poisonous or not with machine learning?

	Agaricus campestris (non-poisonous)	Agaricus californicus (poisonous)
cap color	white	whitish to brown
gill color	pink (young), chocolate brown (mature)	pink (young), chocolate brown (mature)
ring	no	yes
bruises?	no	yes
habitat	grass	forest and grass
odor	mild, mushroomy	antiseptic

Basic mushroom anatomy



For this project, I used the UCI Mushroom Data Set to determine if machine learning could be used to classify the edibility of mushrooms

The data set consists of mushrooms from 2 different families only (Agaricus, Lepiota)

- **23 individual species of mushrooms**

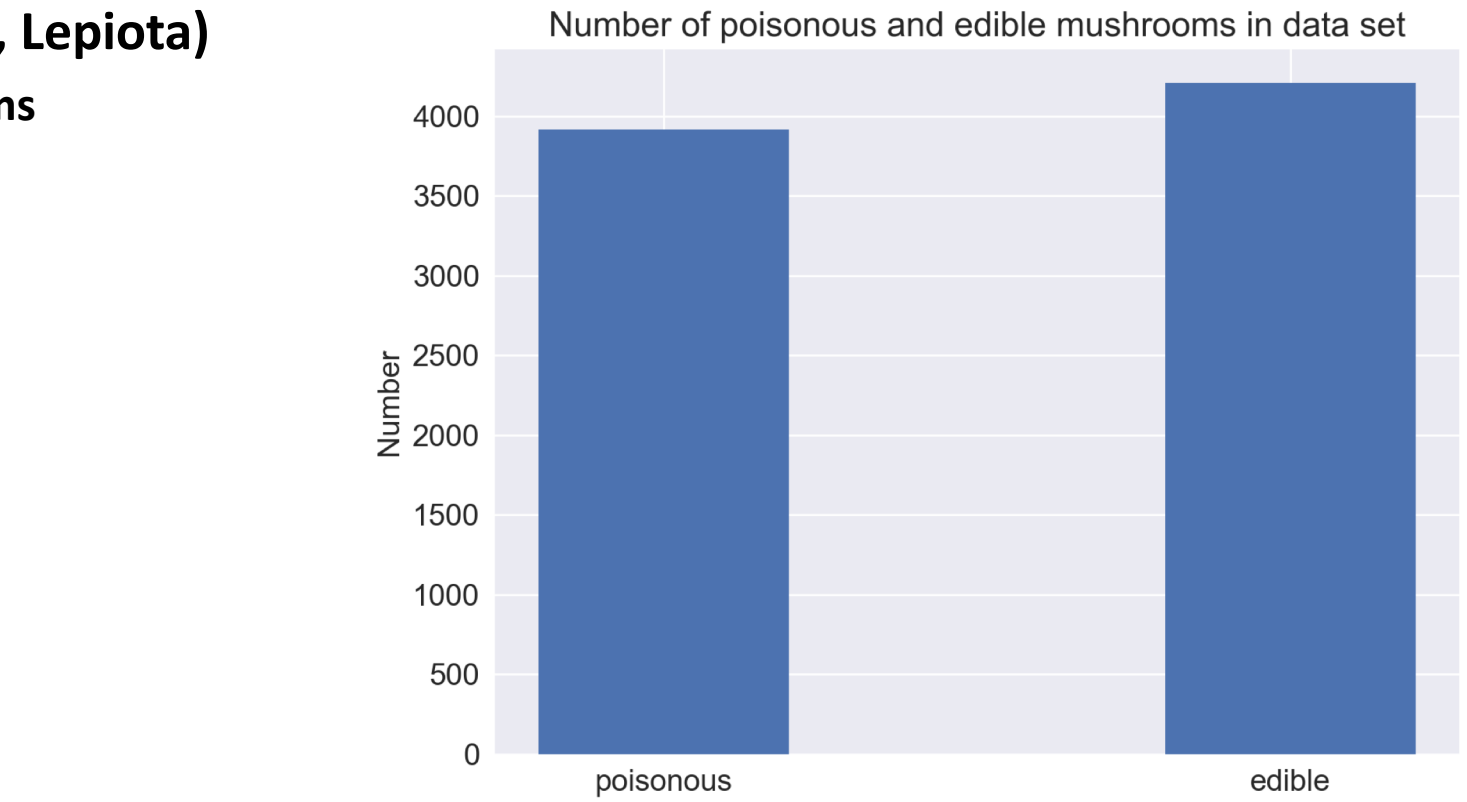
Number of instances = 8124

Number of features = 22

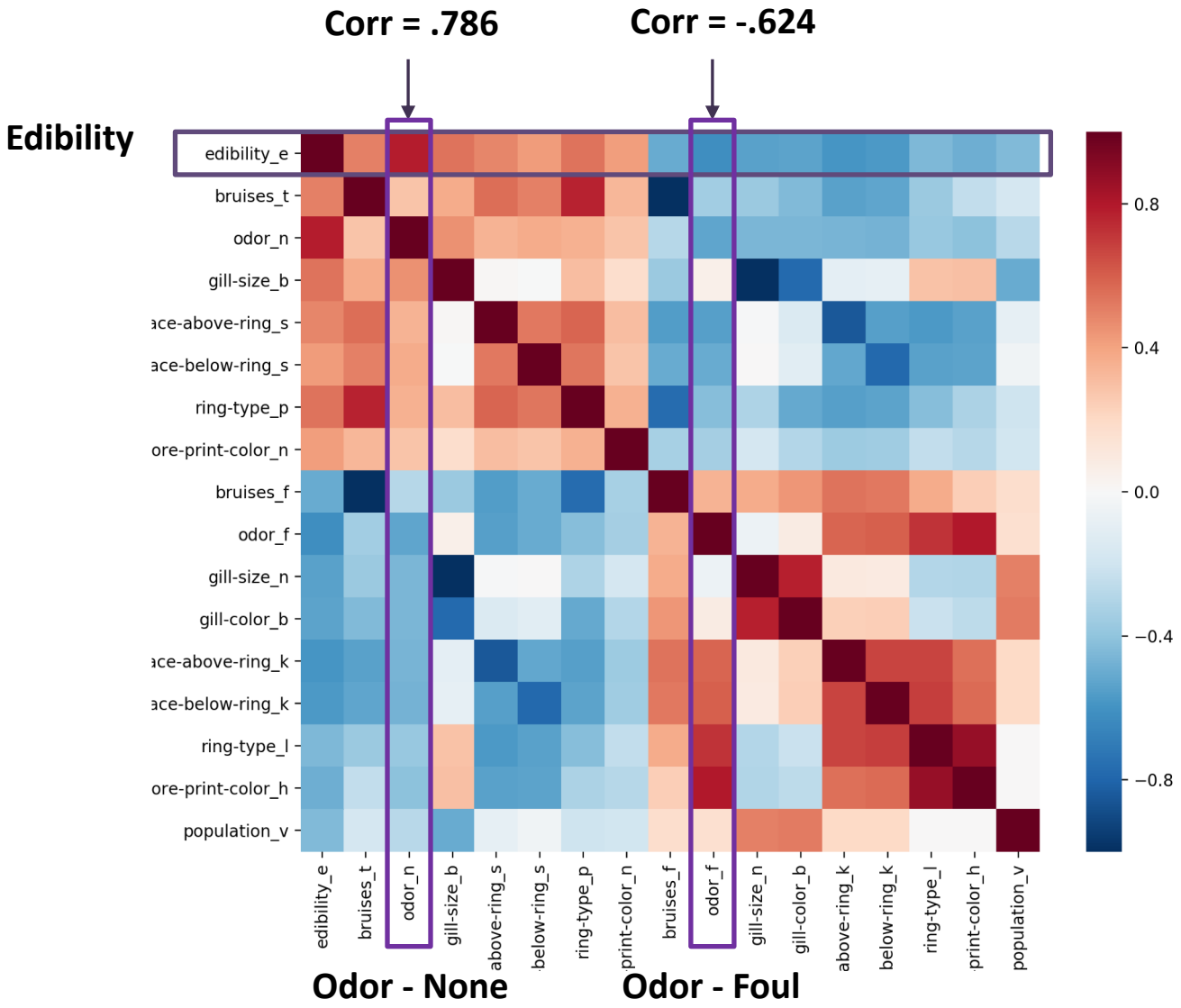
- **All features are categorical**

Target variable is edibility

- **52% edible**
- **48% non-edible**



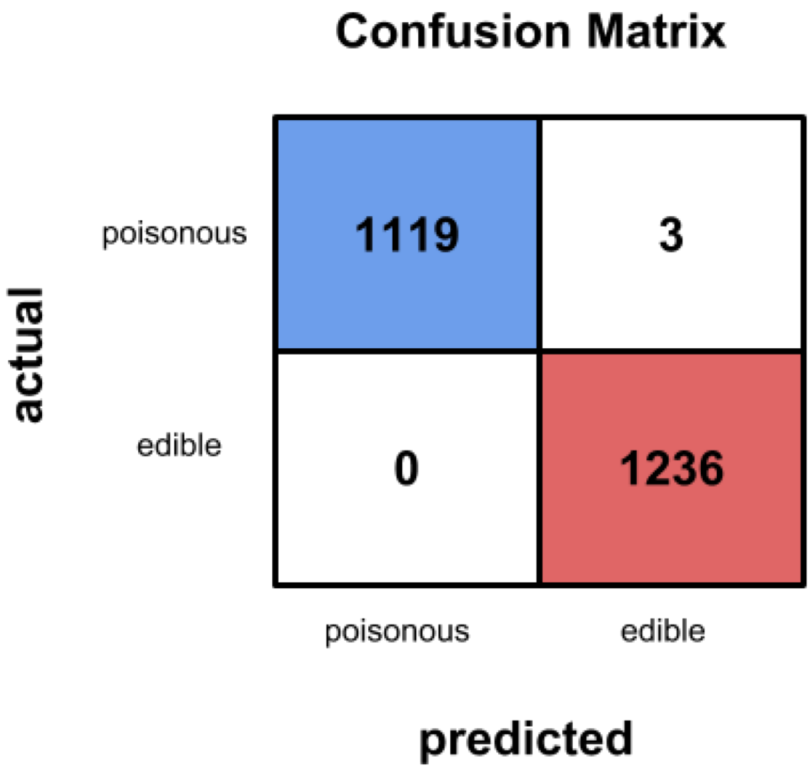
Certain features are strongly correlated with edibility



Baseline model using logistic regression has near perfect accuracy

Logistic Regression Results with default parameters

Model	Precision	Recall	Accuracy	AUC
Logistic Regression	0.998	1.000	0.999	1.000



Which features are most important for predicting edibility?

Logistic regression important coefficients

Feature	β	
odor_n	4.064	Predictive of edibility
odor_l	2.656	
odor_a	2.647	
gill-size_b	1.959	
gill-spacing_c	-1.522	Predictive of poisonousness
stalk-surface-above-ring_k	-1.665	
gill-color_b	-1.799	
gill-size_n	-2.067	
odor_p	-2.253	
stalk-root_b	-2.287	
odor_f	-2.608	
odor_c	-2.638	
spore-print-color_r	-3.331	

For these 2 families of mushrooms, odor is one of the most important features

<u>Odors indicating edibility</u>

n = None, l = anise, a = almond

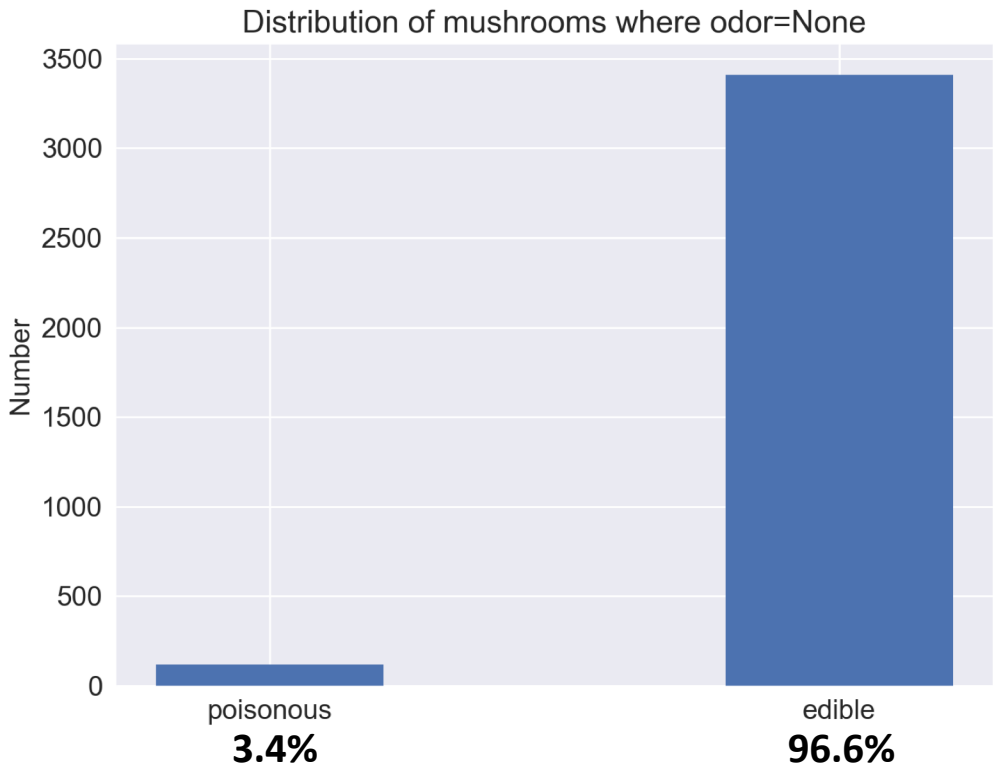
<u>Odors indicating poisonousness</u>

p = pungent, f = foul, c = creosote

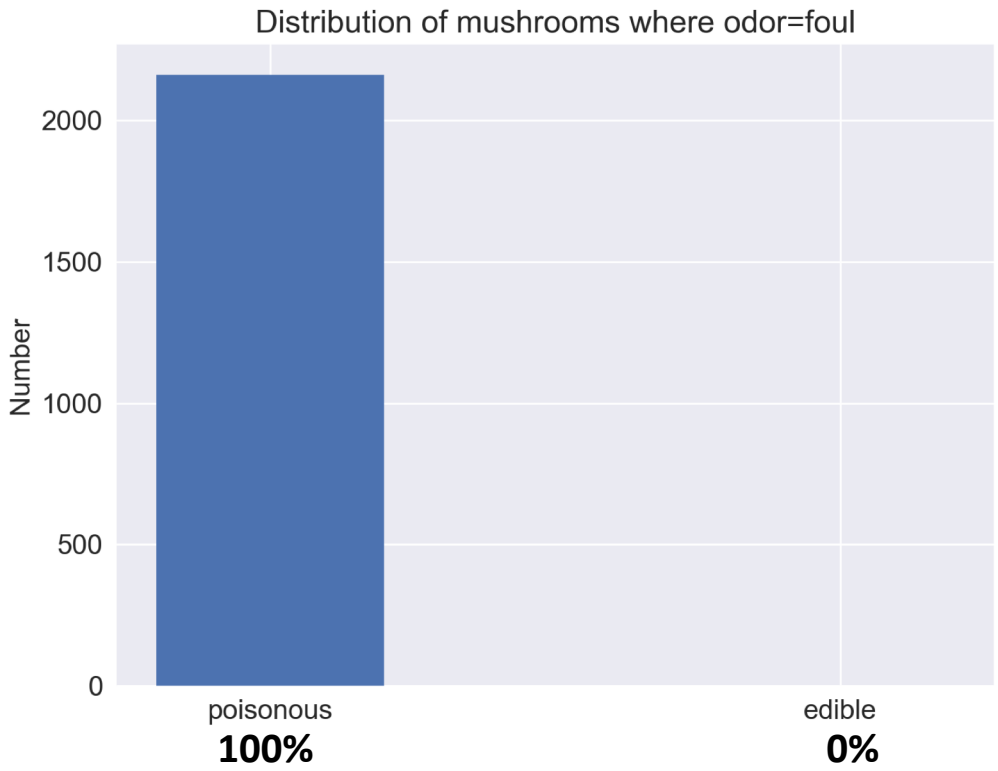
Also, don't eat mushrooms in which the spore print is green

Many mushrooms can be correctly classified just based on two features

β = 4.064

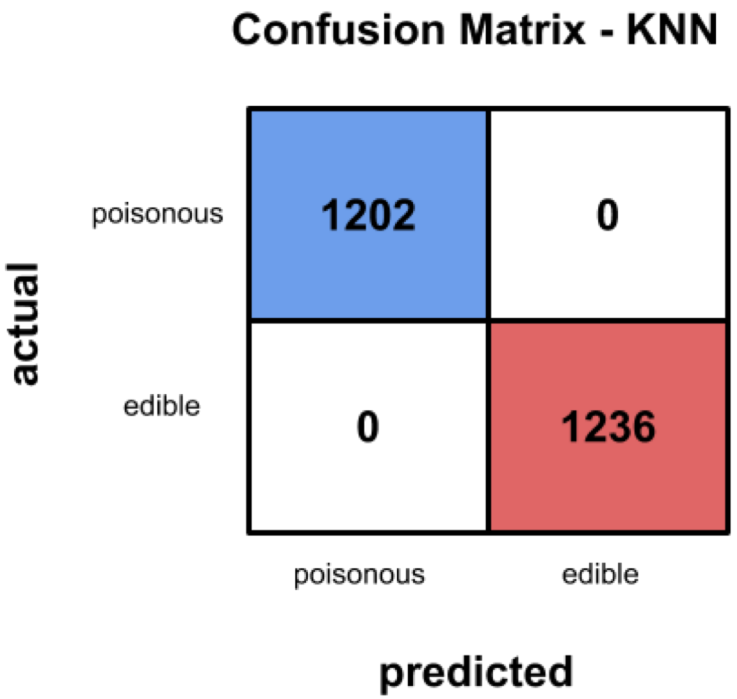


β = -2.608



Other models resulted in perfect accuracy

Model	Precision	Recall	Accuracy	AUC
Logistic Regression	0.998	1.000	0.999	1.000
KNN	1.000	1.000	1.000	1.000
Decision Trees	1.000	1.000	1.000	1.000



What would our model predict for the two mushrooms we saw at the beginning?

Agaricus Campestris (Non-poisonous)

y = 1 (edible)
P(y=1) = 1

Agaricus Californicus (Poisonous)

y = 0 (poisonous)
P(y=0) = 1

In conclusion, we were able to predict mushroom edibility with an accuracy of 1.000

Logistic Regression, KNN, and Decision Trees all did really well for this data set.

Mushroom odor was very predictive of edibility.

Next steps:

- **Explore more mushroom species to see how well we can predict edibility for other mushroom types**
- **See if we can correctly ID mushrooms based on these attributes**
- **Explore using neural networks to classify mushrooms from photos**

Thanks!

Appendix: Mushroom Attribute Information

```
7. Attribute Information: (classes: edible=e, poisonous=p)
     1. cap-shape:                bell=b,conical=c,convex=x,flat=f,
                                  knobbed=k,sunken=s
     2. cap-surface:              fibrous=f,grooves=g,scaly=y,smooth=s
     3. cap-color:                brown=n,buff=b,cinnamon=c,gray=g,green=r,
                                  pink=p,purple=u,red=e,white=w,yellow=y
     4. bruises?:                 bruises=t,no=f
     5. odor:                     almond=a,anise=l,creosote=c,fishy=y,foul=f,
                                  musty=m,none=n,pungent=p,spicy=s
     6. gill-attachment:          attached=a,descending=d,free=f,notched=n
     7. gill-spacing:             close=c,crowded=w,distant=d
     8. gill-size:                broad=b,narrow=n
     9. gill-color:               black=k,brown=n,buff=b,chocolate=h,gray=g,
                                  green=r,orange=o,pink=p,purple=u,red=e,
                                  white=w,yellow=y
    10. stalk-shape:              enlarging=e,tapering=t
    11. stalk-root:               bulbous=b,club=c,cup=u,equal=e,
                                  rhizomorphs=z,rooted=r,missing=?
    12. stalk-surface-above-ring: fibrous=f,scaly=y,silky=k,smooth=s
    13. stalk-surface-below-ring: fibrous=f,scaly=y,silky=k,smooth=s
    14. stalk-color-above-ring:   brown=n,buff=b,cinnamon=c,gray=g,orange=o,
                                  pink=p,red=e,white=w,yellow=y
    15. stalk-color-below-ring:   brown=n,buff=b,cinnamon=c,gray=g,orange=o,
                                  pink=p,red=e,white=w,yellow=y
    16. veil-type:                partial=p,universal=u
    17. veil-color:               brown=n,orange=o,white=w,yellow=y
    18. ring-number:              none=n,one=o,two=t
    19. ring-type:                cobwebby=c,evanescent=e,flaring=f,large=l,
                                  none=n,pendant=p,sheathing=s,zone=z
    20. spore-print-color:        black=k,brown=n,buff=b,chocolate=h,green=r,
                                  orange=o,purple=u,white=w,yellow=y
    21. population:               abundant=a,clustered=c,numerous=n,
                                  scattered=s,several=v,solitary=y
    22. habitat:                  grasses=g,leaves=l,meadows=m,paths=p,
                                  urban=u,waste=w,woods=d
```