



A guide to Mushrooms

Erik Paulson





When are
mushrooms
edible and when
are they poison?



I used the UCI Machine Learning Mushroom Classification dataset to try and discover the answer to this problem

<https://www.kaggle.com/datasets/uciml/mushroom-classification?resource=download>

This dataset includes 23 features of North American mushrooms and 8124 individual data points (individual mushrooms)

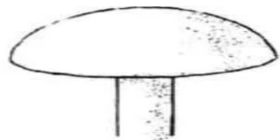
The Problem: Mushrooms are diverse, yet similar

There are 3916 poison mushrooms in our data and some have:

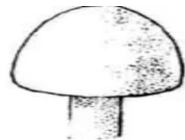
- Convex Caps
- Smooth Surfaces
- Brown Caps ←
- Bruises
- Pungent Odor ←
- Free Gills
- Black Gills

There are 4208 edible mushrooms in our data and some have:

- Convex Caps
- Smooth Surfaces
- Yellow Caps ←
- Bruises
- Almond Odor ←
- Free Gills
- Black Gills



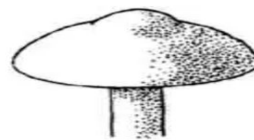
convex



bell-shaped



conical



knobbed





flat



sunken

The Problem: Mushrooms are diverse, yet similar

There are 3916 poison mushrooms in our data and some have:

- Convex Caps
- Smooth Surfaces
- Brown Caps 
- Bruises
- Pungent Odor 
- Free Gills
- Black Gills

POISON

There are 4208 edible mushrooms in our data and some have:

- Convex Caps
- Smooth Surfaces
- Yellow Caps 
- Bruises
- Almond Odor 
- Free Gills
- Black Gills

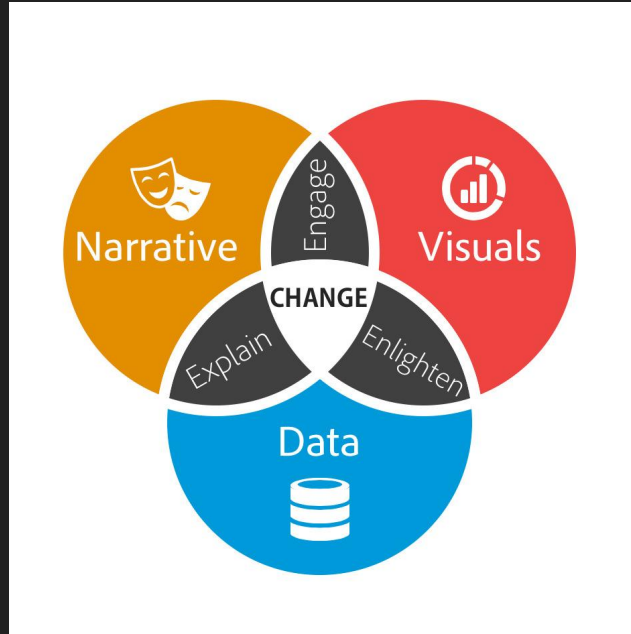
EDIBLE

There are so many trending news articles now such as:

- People are seeking out professional foragers amid fears that grocery stores will run out food
 - New York, Texas, California
 - <https://www.businessinsider.com/foraging-plants-nature-coronavirus-food-2020-5>
- Why foraging is the viral food trend of the moment
 - Canada
 - <https://www.theglobeandmail.com/life/article-why-foraging-is-the-viral-food-trend-of-the-moment/>



Maybe Data Science Can Help



For this Data, I tested a number of different models:

KNN, Logistic Regression, XGBoost:

- Performed very well with an initial score of 1.0

Random Forest must have a depth of at least 7 to obtain a score of 1.0

**We do not want people
wrongfully eating misclassified
poison mushrooms**



Since the models seemed to be cooperating....

XGBoost: Selected model

Accuracy: 1.0

F1: 1.0

Precision: 1.0

Recall: 1.0

(Note 1.0 is highest score)

**We do not want people
wrongfully eating misclassified
poison mushrooms**



FP



FN

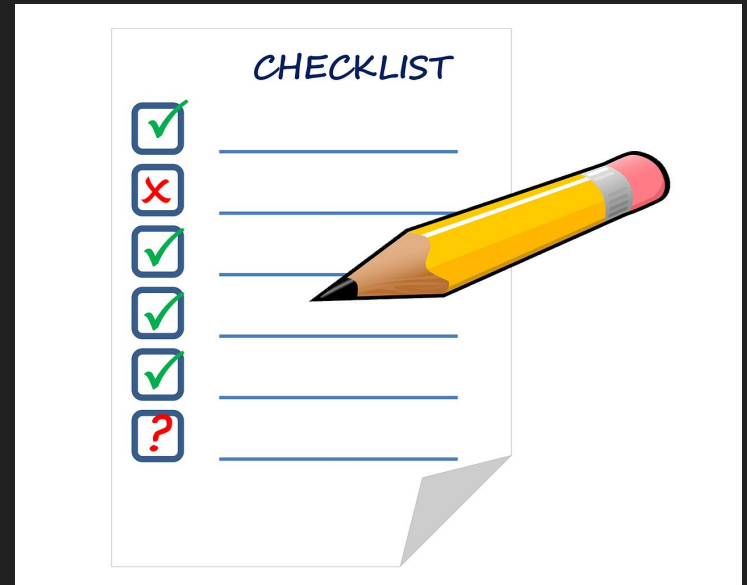
FN

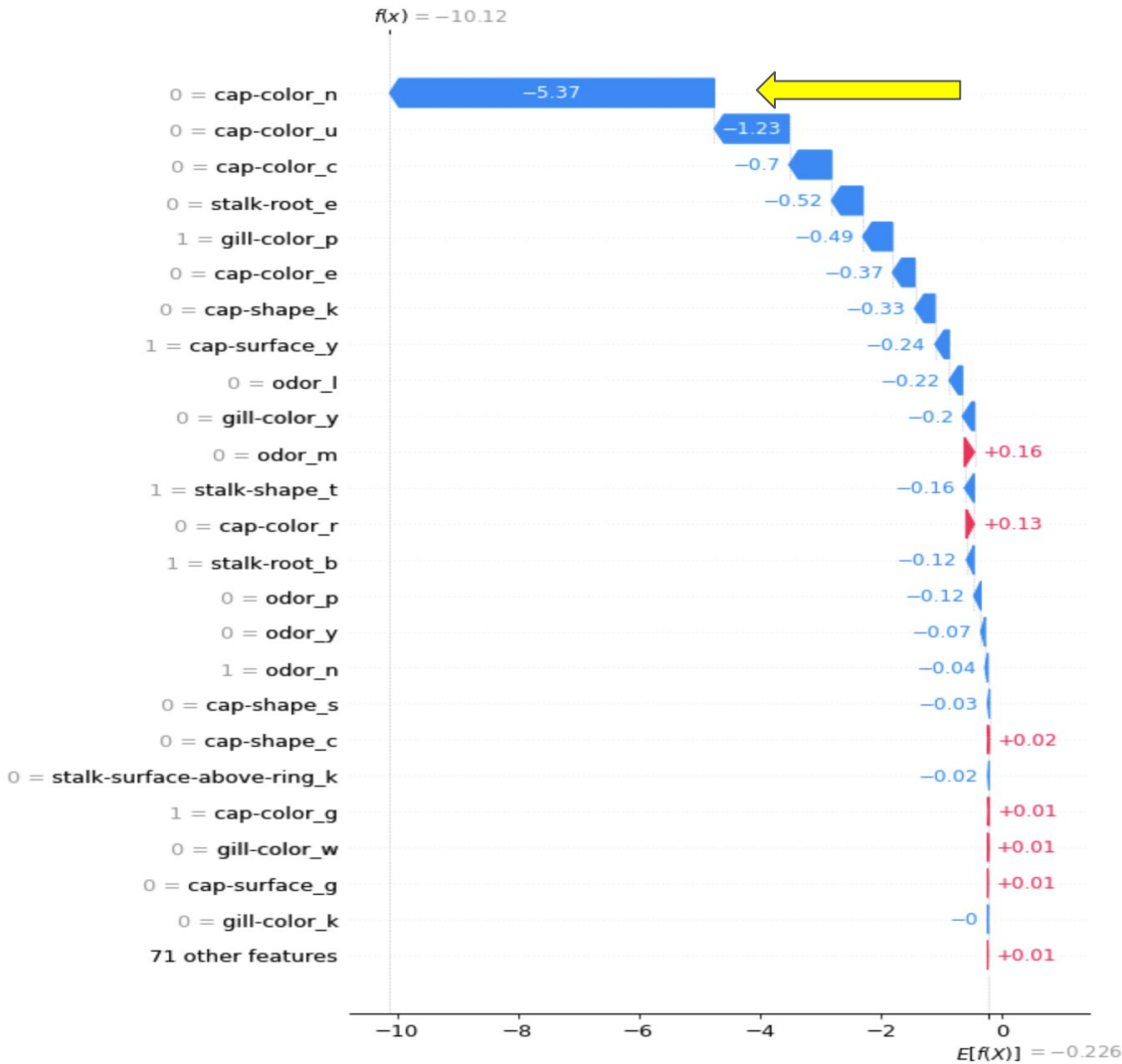




Can the models help find a “golden rule for mushrooms”?

Unfortunately, it doesn't seem like there is an obvious rule...however we can help pinpoint which features are the best to look out for in order to get an accurate prediction





Using SHAP values we can see the importance of different features for the model

The top features of importance are:

1. Cap-color_n (Brown)
2. Cap-color_u (Purple)
3. Cap-color_c (Cinnamon)
4. Stalk-root_e (Equal)
5. Gill-color_p (Pink)

Future Work

The next step to help everyday foragers:

- Potentially create an app that utilizes our model based on inputted features for foragers to use



The next step scientifically:

- Give our model to biologists and researchers to use when a new species of mushroom is discovered...
- obviously additional tests must be done before simply eating the mushroom, but our models could point in correct directions

Conclusion

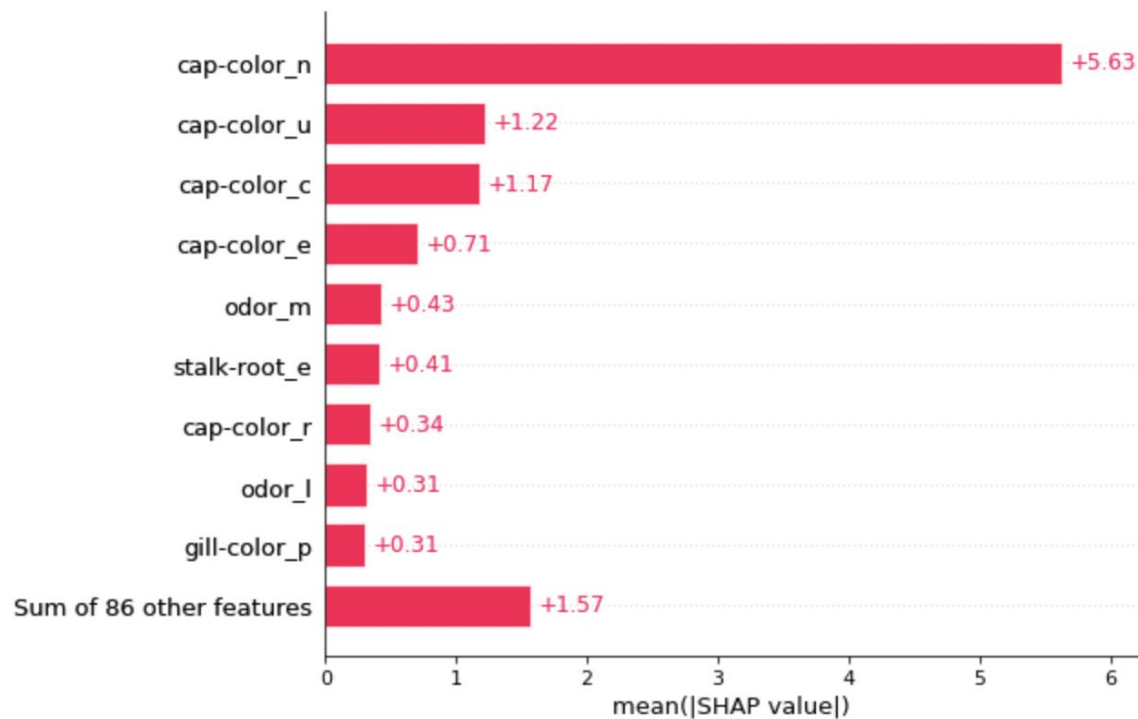
Our goal is to use advanced models to help foragers to correctly gather safe to eat mushrooms and avoid poisonous ones

There is no easy way to do this with a standard rule, only with a trained guide or... using one of our accurate models.

There may be further uses for these models! Such as in a personal use app, or as a model for researchers to follow when evaluating new species.

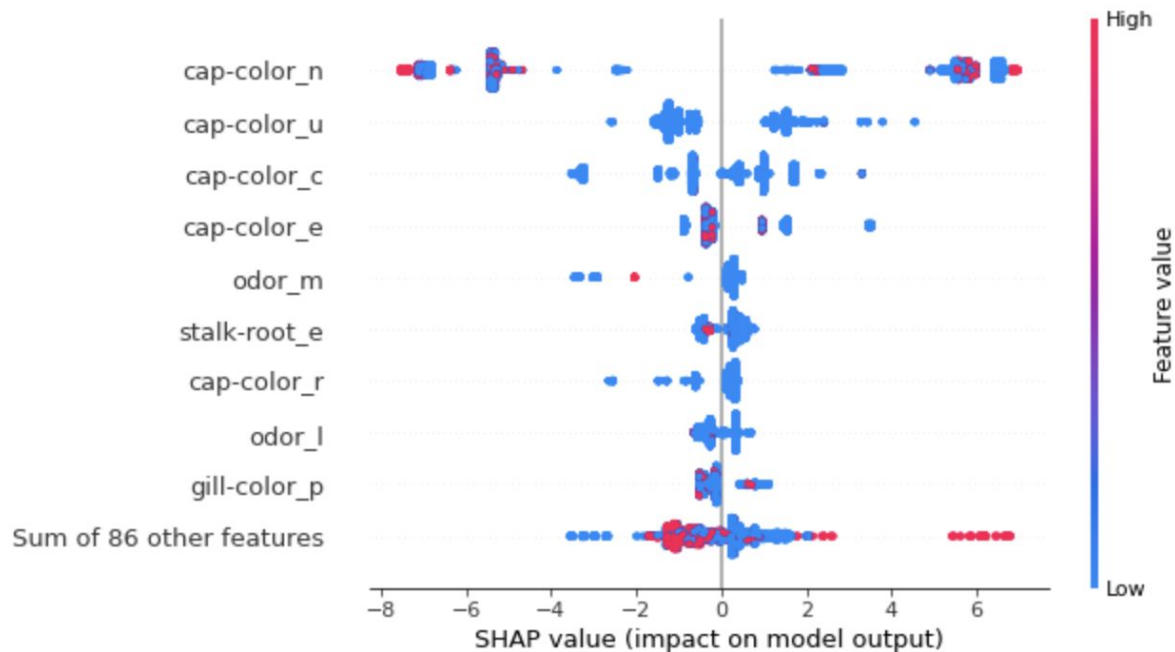
Index

```
[94]: shap.plots.bar(shap_values)
```

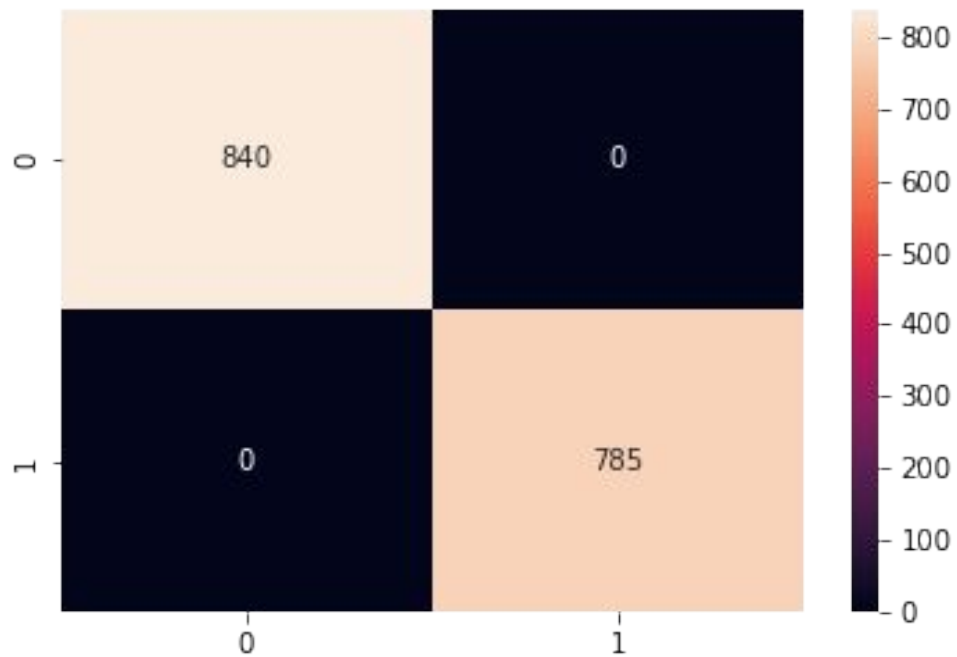


Index

```
[96]: shap.plots.beeswarm(shap_values)
```



Index



Index

