Assignment 8

Mushroom Stew

Introduction

* Analyze the mushroom dataset to determine which mushrooms are best to add to a stew
* Extract information like taste, toxicity, and visual appeal of a mushroom

Dataset Analysis

Dataset features:

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

* Labels: 1=definitely edible, 0=definitely poisonous or unknown edibility
  + For our purposes, we’ll lump unknown into poisonous instead of treating it like its own class

Inputs

* [UCI Mushroom Dataset](https://archive.ics.uci.edu/ml/datasets/mushroom)

Proposed Libraries

* Pandas, sklearn, matplotlib, plotly, scipy, seaborn
* All libraries will come from pip
* <https://pypi.org/project/pandas/>
* <https://pypi.org/project/scikit-learn/>
* <https://pypi.org/project/matplotlib/>
* <https://pypi.org/project/plotly/>
* <https://pypi.org/project/scipy/>
* <https://pypi.org/project/seaborn/>

Proposed Solution

* First, use group consensus to categorize qualitative features, like smell and visual appeal, into “good” and “bad” classes
* Drop all other features which do not contribute to the overall “stewability” of a mushroom
* Standardize data and fit multiple classification models
* Compare all of the classifiers and select the most accurate one

Proposed Outputs

* A list of each mushroom and its stewability score. From that, a selection from the top mushrooms will make it onto our final recipe.
* A list of the best mushrooms that are also toxic, for our Widower’s stew.

Proposed Visualization

* Model performances
* Feature importance
* Pictures of winning mushrooms

Conclusions

* A tasty mushroom stew recipe (that won’t kill you).
* Recommended wine pairings for the stew
* A Widower’s stew that tastes and looks delicious but will definitely kill you.