Mushroom Type Prediction

Abstract:

Mushroom hunting (otherwise known as "shrooming") is enjoying new peaks in popularity. In this case study we will learn which features spell certain death and which are most palatable in this dataset of mushroom characteristics.

Problem Statement:

Use decision tree algorithm to predict whether a mushroom sample is poisonous or edible

Dataset Information:

This dataset includes descriptions of hypothetical samples corresponding to 23 species of gilled mushrooms in the Agaricus and Lepiota Family Mushroom.Each species is identified as definitely edible, definitely poisonous, or of unknown edibility and not recommended. This latter class was combined with the poisonous one. The Guide clearly states that there is no simple rule for determining the edibility of a mushroom; no rule like "leaflets three, let it be'' for Poisonous Oak and Ivy.

Variable Description:

Column	Description
cap-shape	bell=b,conical=c,convex=x,flat=f, knobbed=k,sunken=s
cap-surface	fibrous=f,grooves=g,scaly=y,smooth=s
cap-color	brown=n,buff=b,cinnamon=c,gray=g,green=r,pink=p,purple=u,red=e,white=w,yellow=y
bruises	bruises=t,no=f
odor	almond=a,anise=l,creosote=c,fishy=y,foul=f,musty=m,none=n,pungent=p,spicy=s
gill-attachment	attached=a,descending=d,free=f,notched=n
gill-spacing	close=c,crowded=w,distant=d

gill-size	broad=b,narrow=n
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
stalk-shape	enlarging=e,tapering=t
stalk-root	bulbous=b,club=c,cup=u,equal=e,rhizomorphs=z,rooted=r, missing=?
stalk-surface-above-ring	fibrous=f,scaly=y,silky=k,smooth=s
stalk-surface-below-ring	fibrous=f,scaly=y,silky=k,smooth=s
stalk-color-above-ring	brown=n,buff=b,cinnamon=c,gray=g,orange=o,pink=p,red=e,white=w,yellow=y
stalk-color-below-ring	brown=n,buff=b,cinnamon=c,gray=g,orange=o,pink=p,red=e,white=w,yellow=y
veil-type	partial=p,universal=u
veil-color	brown=n,orange=o,white=w,yellow=y
ring-number	none=n,one=o,two=t
ring-type	cobwebby=c,evanescent=e,flaring=f,large=l,none=n,pendant=p,sheathing=s,zone=z
spore-print-color	black=k,brown=n,buff=b,chocolate=h,green=r,orange=o,purple=u,white=w,yellow=y
population	abundant=a,clustered=c,numerous=n,scattered=s,several=v,solitary=y
habitat	grasses=g,leaves=l,meadows=m,paths=p,urban=u,waste=w,woods=d
class	edible=e, poisonous=p

Scope:

- Exploring insights about the unfortunate trip from the existing data
- Data Pre-processing
- Training Decision Tree algorithms to predict the whether mushroom is edible or poisonous
- Evaluating the model with various metrics like Accuracy, AUC ROC, Precision, etc. and improve the score using statistical analysis over time

Learning Outcome:

The students will get a better understanding of how the variables are linked to each other and build a decision tree classification model. They will also learn about various performance measures of classification models and should be able to improve these scores by taking the necessary step.