Mushroom Classification

Detailed Project Report

- Varun Salunkhe
- Sourabh Hawale
- Saurabh Jumnalkar

iNeuron

PROJECT DETAIL

Project Title	Mushroom Classification
Technology	Machine Learning Technology
Domain	Agriculture
Project Difficulty level	Intermediate
Programming Language Used	Python
Tools Used	Jupyter Notebook, Vscode, GitHub,MongoDB compass

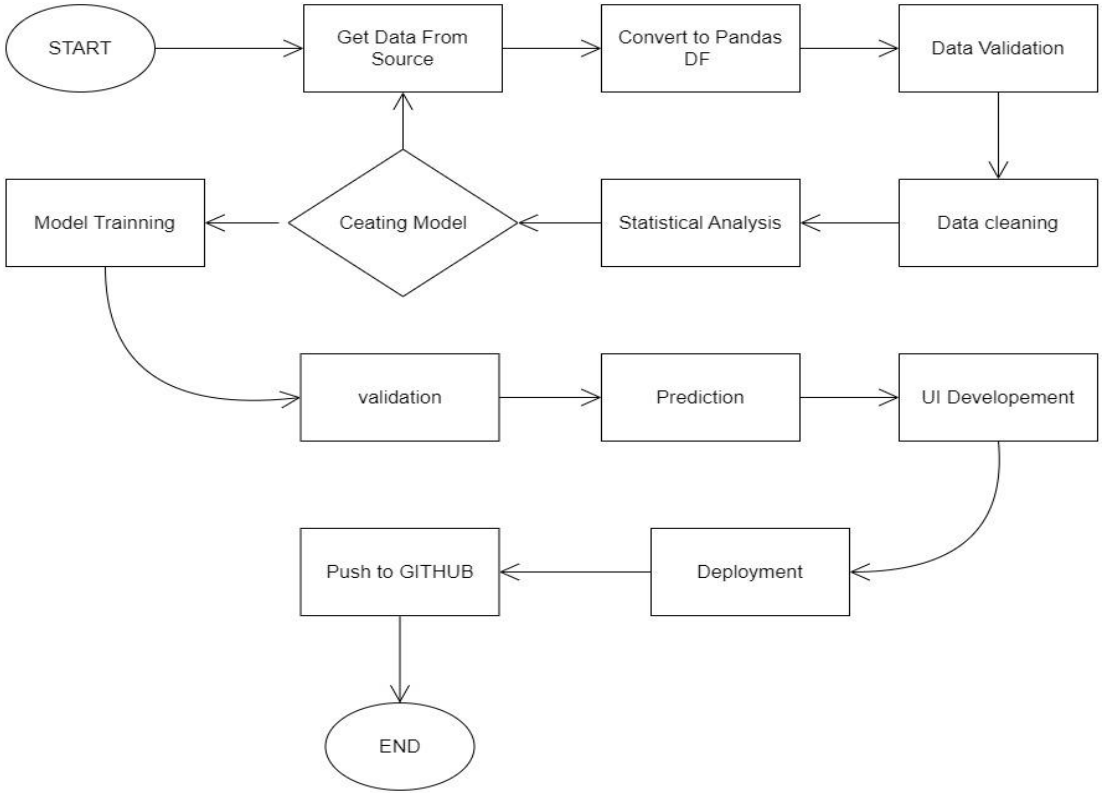
OBJECTIVE

- The main goal is to predict which mushroom is poisonous & which is edible.

PROBLEM STATEMENT

- The Audubon Society Field Guide to North American Mushrooms contains descriptions of hypothetical samples corresponding to 23 species of gilled mushrooms in the Agaricus and Lepiota Family Mushroom (1981). Each species is labelled as either definitely edible, definitely poisonous, or maybe edible but not recommended. This last category was merged with the toxic category. The Guide asserts unequivocally that there is no simple rule for judging a mushroom's edibility, such as "leaflets three, leave it be" for Poisonous Oak and Ivy.

ARCHITECTURE



DATASET INFORMATION

Attribute Information: (classes: edible=e, poisonous=p)

- 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

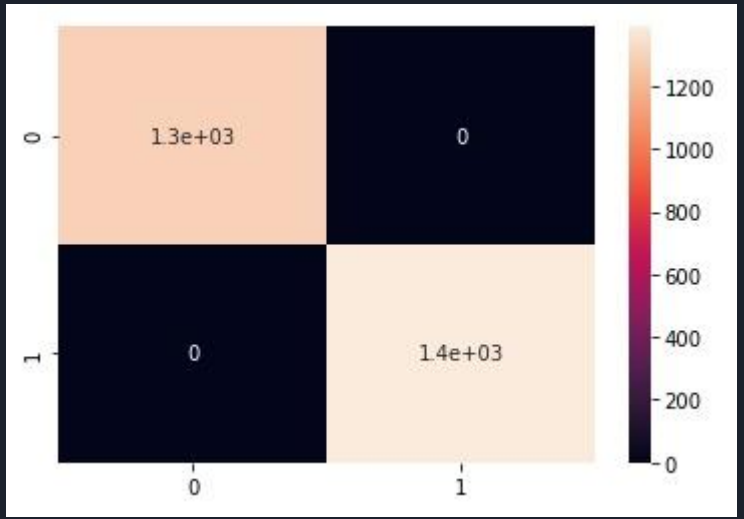
KEY PERFORMANCE INDICATOR (KPI)

- By using the Wrapper method and Filter method, the Key Attributes that contributed to the better classification of mushrooms are identified.
- The attributes that have been found to be the best ones from both the attribute selection methods are compared.
- It is found that both the attribute selection methods almost gave the same results as the output. Hence by using these attributes as the key attributes, there will be better accuracy in the classification of mushrooms as edible or poisonous.
- The key attributes were also found to have good Precision, Recall, and F-Measure values.

PERFORMANCE

- Precision score

- A much better way to evaluate the performance of a classifier is to look at the confusion matrix. The general idea is to count the number of times instances of class A are classified as class B.
- For example, to know the number of times the classifier confused images of 5s with 3s, you would look in the 5th row and 3rd column of the confusion matrix.



CONCLUSION

This project is about the methods of pre-processing, steps to identify the key attributes that help in the better classification of edible and poisonous mushrooms, and also a comparison between the attribute selection methods in order to find whether both methods produce the same output.

THANK YOU