High-Level Design (HLD)

Mushroom Classification

Revision Number: 2.0

Last Date of revision: 08/12/2022

Document Version control

Date issued	Version	Descriptions	Author
20/11/2022	1.0	Abstract, Introduction	Varun Salunkhe
30/11/2022	1.1	General Description	Saurabh Jumnalkar
05/12/2022	1.2	Design Details	Saurabh Jumnalkar
08/12/2022	1.3	KPLs, Conclusion	Sourabh Hawale

Contents

Document Version Control

Abstract

Abstract

Mushroom is found to be one of the best nutritional foods with high proteins, vitamins, and minerals. It contains antioxidants that prevent people from heart disease and cancer. Around 45000 species of mushroom are found to be existing in worldwide. Among these, only some of the mushroom varieties were found to be edible. Some of them are really dangerous to consume. In order to distinguish between the edible and poisonous mushrooms in the mushroom dataset which was obtained from UCI Machine Learning Repository, some data mining techniques are used. Weka is a data mining tool with various machine learning algorithms that can pre-process, analyze, classify, visualize and predict the given data. Thus, to select the attributes that help better classify mushrooms, the Wrapper method and Filter method in Weka is used to identify the best attributes for the classification. The attributes 'odor' and 'spore_print_color' were chosen to be the best ones that contributed to the better classification of edible and poisonous mushrooms. After identifying the key attributes, classification is performed, a decision tree is constructed based on those attributes, and its Precision, Recall, and F-Measure values are analyzed.

1. Introduction

1.1 Why this High-Level Design?

The purpose of this High-Level Design (HLD) Document is to add the necessary detail to the current project description to represent a suitable model for coding. This document is also intended to help detect contradictions prior to coding, and can be used as a reference manual for how the modules interact at a high level.

The HLD will:

- Present all of the design aspects and define them in detail Describe the user interface being implemented
- Describe the hardware and software interfaces
- Describe the performance requirements
- Include design features and the architecture of the project
- List and describe the non-functional attributes like:
    - Security
    - Reliability
    - Maintainability
    - Portability
    - Reusability
    - Application compatibility
    - Resource utilization
    - Serviceability

1.2 Scope

The HLD documentation presents the structure of the system, such as the database architecture, application architecture (layers), application flow (Navigation), and technology architecture. The HLD uses non-technical to mildly-technical terms which should be understandable to the administrators of the system.

2. General Description

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

2.2 Data Requirements

Although this dataset was originally contributed to the UCI Machine Learning repository nearly 30 years ago, mushroom hunting (otherwise known as "shrooming") is enjoying new peaks in popularity. Learn which features spell certain death and which are most palatable in this dataset of mushroom characteristics. And how certain can your model be?

2.3 Data Content

This dataset includes descriptions of hypothetical samples corresponding to 23 species of gilled mushrooms in the Agaricus and Lepiota Family Mushroom drawn from The Audubon Society Field Guide to North American Mushrooms (1981). 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.

2.4 About this File

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,punge nt=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,yell ow=y
- stalk-color-below-ring: brown=n,buff=b,cinnamon=c,gray=g,orange=o,pink=p,red=e,white=w,yell ow=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

2.5 Tools Used

2.5.1 Hardware Requirements

- All Operating System
- Min 500MB ram
- I3 intel

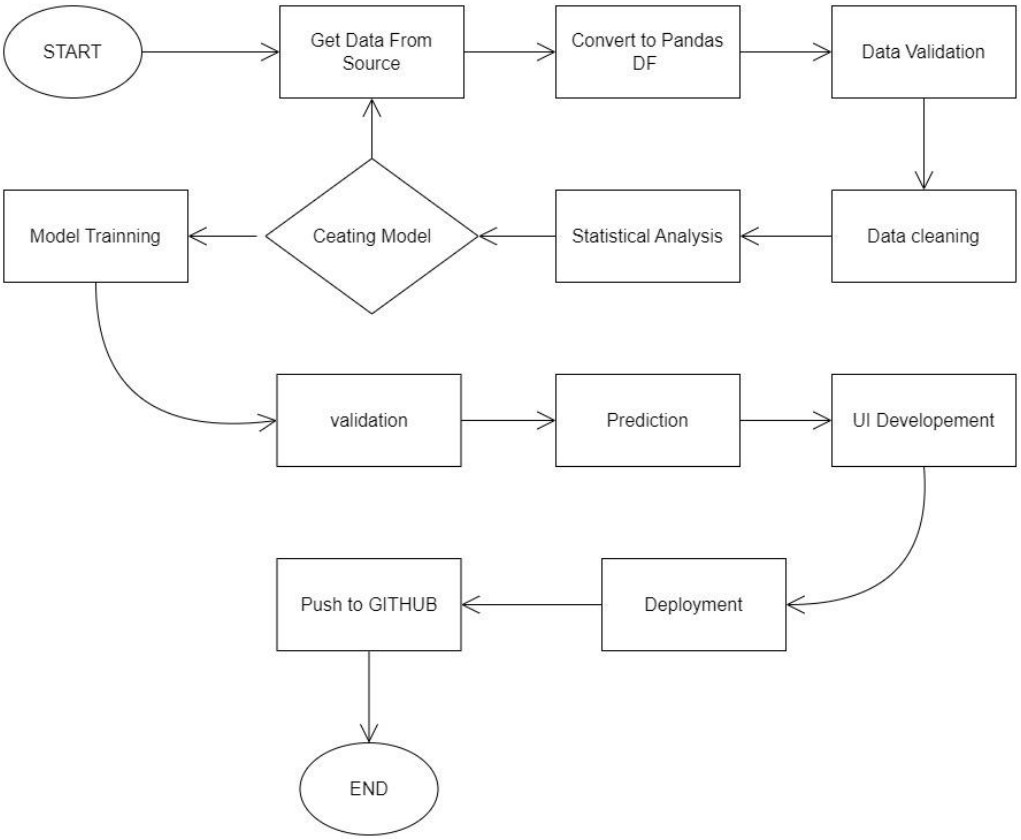
2.5.2 Software Requirements

- Python
- Flask Framework
- MongoDB

- Vs Code
- Numpy
- Pandas
- Matplotlib
- Jupyter

3. Design Details

3.1 Process Flow



3.2 Event log

We should be able to log every activity done by the user.

- The System identifies at what step logging required
- The System should be able to log each and every system flow.
- Developers can choose logging methods. You can choose database logging/File logging as well.
- The system should not be hung even after using so many loggings. Logging is just because we can easily debug issues so logging is mandatory to do.

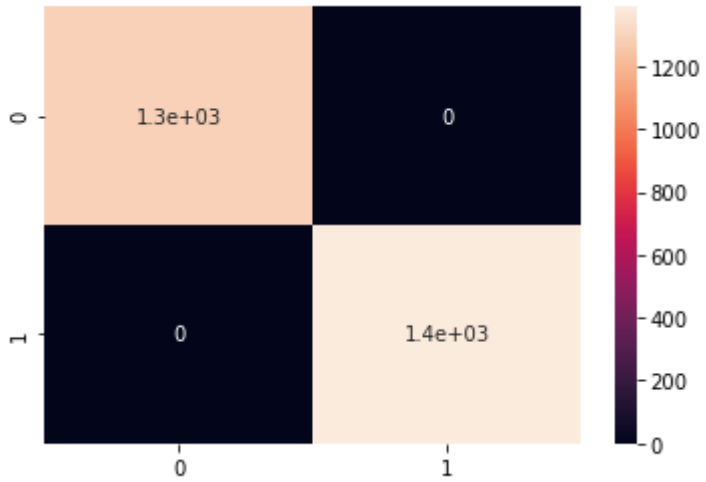
3.3 Error Handling

In Python, A traceback is a report containing the function calls made in your code at a specific point i.e when you get an error it is recommended that

you should trace it backward (traceback). Whenever the code gets an exception, the traceback will give the information about what went wrong in the code. Python traceback contains great information that can help you find what is going wrong in the code. These tracebacks can look a little wearisome, but once you break them down to see what it's trying to show you, they can be very helpful.

3.4 Performance

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.



4. Dashboards

4.1 KPLs

Thus, 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.

5. 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.