High Level Design (HLD)

Mushroom Classification

Revision Number:
Last date of revision:

Document Version Control

Date Issued	Version	Description	Author
11-11-2021	1	Initial HLD-V1.0	Oviyashri B

Contents

Document Version Control 2

Abstract 4

1 Introduction 5

Why this High-Level Design Document---------- 5
Scope 5
Definitions 5

2 General Description 6

Product Perspective 6
Problem Statement 6
Proposed Solution 6
Further Improvements 6
Data Requirements 7
Tools Used 8
Constraints 9
Assumptions 9

3 Design Details 9

Process Flow 9
3.1.1 Model Training and Evaluation ----------10
3.1.2 Deployment Process 10
Event log 11
Error Handling 11

4 Performance 12

Reusability 12
Application Compatibility 12
Resource Utilization 12
Deployment 12

5 Conclusion 13
6 References 13

Abstract

Mushrooms can be found extensively in a variety of natural environments and visual identification of mushroom species is well established. Some mushrooms are known because of their nutritional and therapeutic properties. Some species are known all over the world because of their toxicity that causes fatal accidents every year mainly due to misidentification. Some of the edible mushrooms are *Ganodeíma spp, Canthaíellus spp, Agaricusspp, Pleuíotus spp, Russula spp, Auricularia spp and ľermitomyces spp;* but the ornamentals are the beautifullyringed*Micropoíous spp. Amanita spp, Lepiota cristata, Lepiota brunneoincarnata and Inocybe asterospoía, Coprinusspp* are among the most important species responsible for mushroom poisoning. Morphological and chemical analyses for mushrooms are occasionally required in forensic science practice. In this work, the characteristics of the representative toxic mushrooms and some chemical methods for their toxins are presented. Mushrooms are identified traditionally by their appearance, taste, colour, odour, presence of scales etc

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 may be 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. The classical machine learning tasks like Data Exploration, Data Cleaning, Feature Engineering, Model Building and Model Testing.

1 Introduction

Why this High-Level Design Document?

The purpose of this High Level Design (HLD) Document is to add the necessary details 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 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 feature and the architecture of the project
- List and describe the non-functional attribute like:
    - Security
    - Reliability
    - Maintainability
    - Portability
    - Reusability
    - Application compatibility
    - Resource utilization
    - Serviceability

Scope

The HLD document 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.

Definitions

- MP – Mushroom Prediction

2 General Description

Product Perspective

The Mushroom prediction solution system is a data science-based machine learning model which help us to detect the edible and poisonous mushrooms take necessary action.

Problem Statement

To create an AI solution for predicting mushrooms and to implement the following use cases.

- To detect Mushroom prediction and its type in edible.
- To detect Mushroom prediction and its type in poison.

Proposed Solution

The solution proposed here is a data science model based on machine learning can be implemented to perform above mention use cases. In first use case , we will take input from edible and also poisonous, check our solution whether it is performing or not in right way.

Further Improvements

The Mushroom prediction solution can be added with more use cases in agriculture domain, to give one step extra confirmation of health to those people who intake mushrooms.

Data Requirements

Data requirement completely depend on our problem statement.

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

Tools used

Python programming language and frameworks such as NumPy, Pandas, Scikit-learn,Matplotlib, Flask, Streamlit etc are used to build the whole model.


- Virtual Studio Code is used as IDE
- For visualization of the plots, Matplotlib are used.
- Heroku and streamlit are used for deployment of the model.
- Python, Streamlit are used for backend development
- Github is used as Version Control System.
- DVC is used for Data Version Control

Constraints

The MP solution system must be correct enough that it not mislead any report and as automated as possible and users should not be required to know any of the workings.

Assumptions

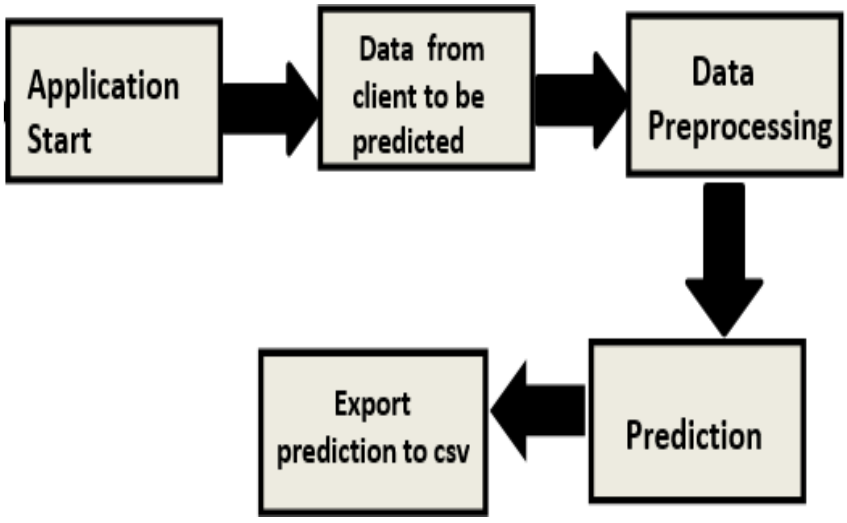
The main objective of the project is to implement the use cases as previously mentioned for new dataset that comes through agriculture which has this solution install in their campus to capture mushroom reports.

3 Design Details

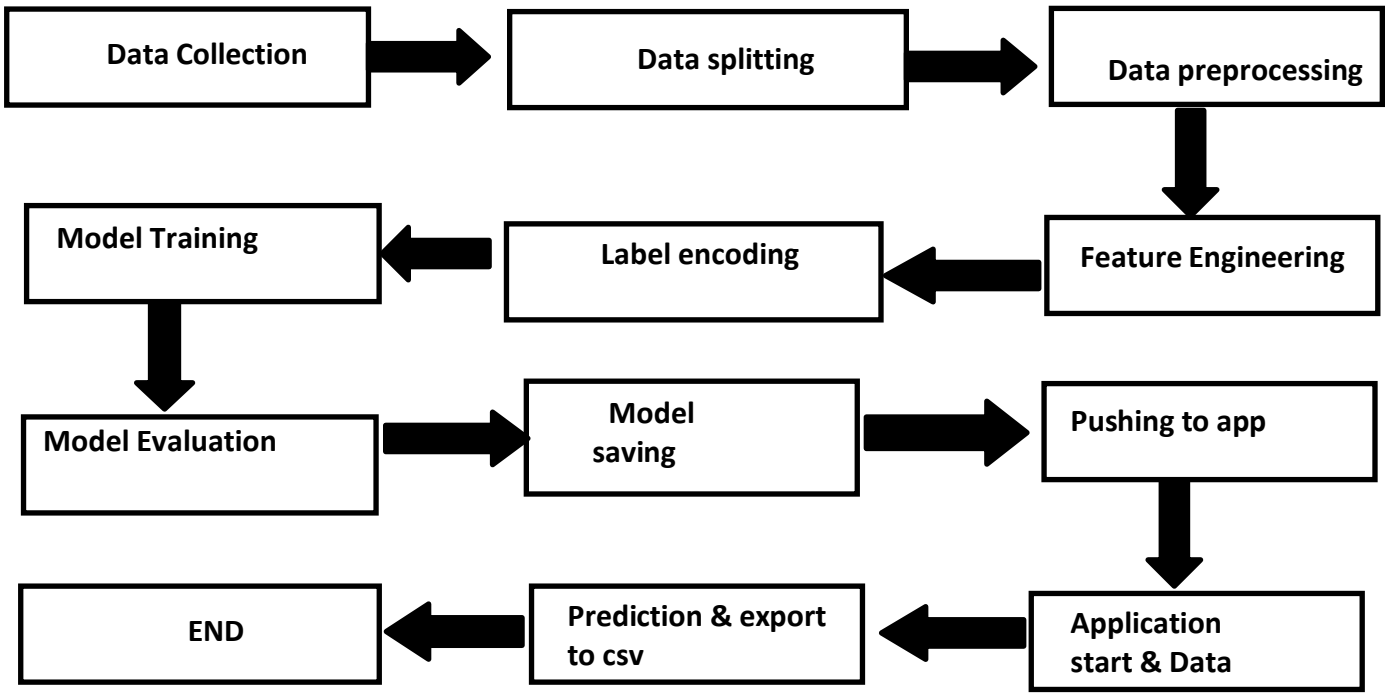
Process Flow

For mushroom prediction, we will use machine learning base model. Below is the processflow diagram is as shown below

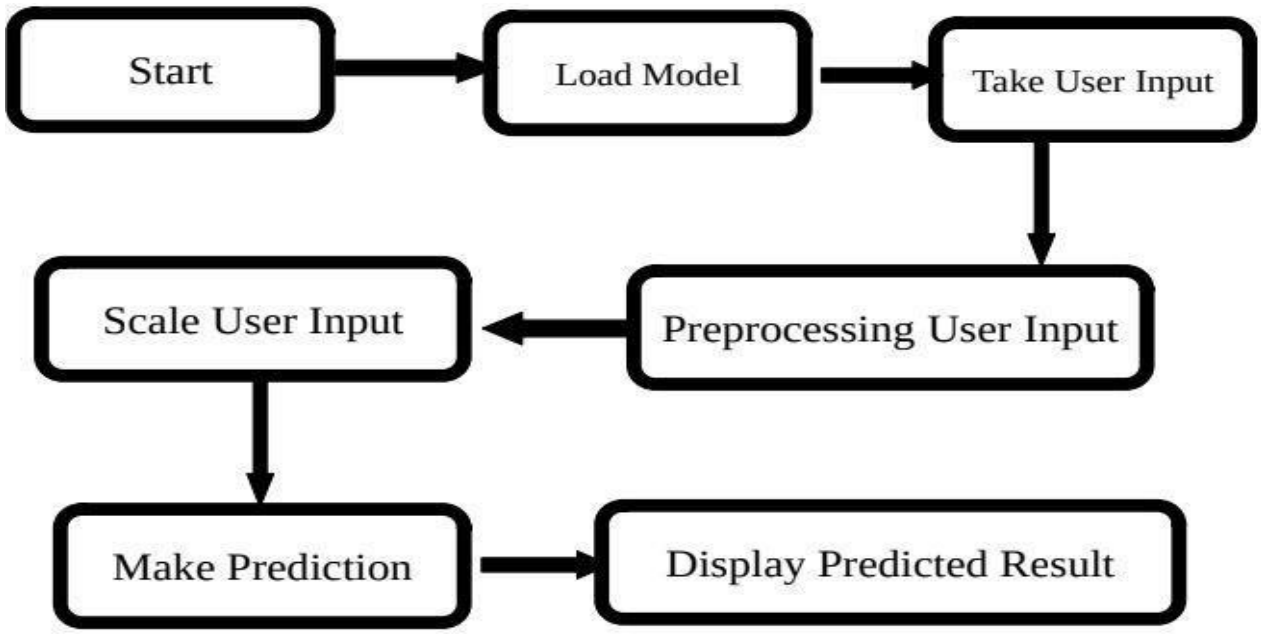
Proposed methodology



Model Training and Evaluation



Deployment Process



Event log

The system should log every event so that the user will know what process is running internally.

Initial Step-By-Step Description:

1. The System identifies at what step logging required.
2. The System should be able to log each and every system flow.
3. Developer can choose logging method. You can choose database logging/ File logging s well.
4. System should not hang even after using so many loggings. Logging just because we can easily debug issues so logging is mandatory to do.

Error Handling

Should errors be encountered, an explanation will be displayed as to what went wrong? An error will be defined as anything that falls outside the normal and intended usage.

4 Performance

The machine learning based Mushroom Prediction solution will used for detection of poison one. So that necessary action will be taken ASP. Also model retraining is very important toimprove performance.

Reusability

The code written and the components used should have the ability to be reused with no problems.

Application Compatibility

The different components for this project will be using python as an interface between them. Each component will have its own task to perform, and it is the job of the Python to ensure proper transfer of information.

Resource utilization

When any task is performed, it will likely use all the processing power available until that function is finished.

Deployment

5 Conclusion

Mushroom prediction was done with the help of machine learning with high accuracy and large amount of data can be predicted at a time. This solution should be as accurate as possible, so that chances of misleading reports will be taken good care of.

6 References

Kaggle For Data Set

URL: https://www.kaggle.com/uciml/mushroom-classification