Mushroom Classifier (HLD)

High Level Document (HLD)

Document Version Control

Date issued	Version	Description	Author
7th November 2023	0.1	Initial HLD – V0.1	Suraj Mali

Abstract

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 aim of the project is to predict which mushroom is poisonous & which is edible by building Machine Learning model with tasks of Data Exploration, Data Cleaning, Feature Engineering, Model Building and Model Testing on different models.

1. Introduction

1.1 Why this High-Level Design Document?

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:
    - o Reliability
    - o Security
    - o Maintainability
    - o Portability
    - o Reusability
    - o Application compatibility
    - o Resource utilization
    - o 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.

1.3 Definitions

Term	Description
Database	**Collection of all information monitored by this system**
IDE	**Integrated Development Environment**
HTML	**Hypertext Markup Language**
ML	**Machine Learning**

2. *General Description*

2.1 Product Perspective

The mushroom classifier is a machine learning-based classifier model which help us to predict the type of mushroom based on its feature.

2.2 Problem Statement

To develop an interface to predict the type of mushroom by choosing categories of its different features like cap shape, cap surface, odor, gill color, population, habitat, etc.

2.3 Proposed Solution

The solution proposed here is to output the type of mushroom being edible or poisonous based on categories chosen on its different features. After result is obtained, we can notify if it is safe to eat or not.

2.4 Further Improvements

The suggested solution initially focuses on selecting categories within the features for classification. However, it can be further enhanced by developing a model using image-based classification through Deep Learning computer vision techniques. This approach would enable the model to classify mushrooms based on the input image itself, expanding its capabilities beyond feature-based classification.

2.5 Technical Requirements

The solution that is proposed can be implemented as a cloud-based solution or as an application hosted on an internet browser as a local server in a local machine. To access this application, the following requirements are proposed:

- Any web browser
- An internet connection (if hosted on cloud)

2.6 Data Requirements

The data for the mushroom classifier includes obtaining a CSV file from Kaggle. The dataset contains information on mushroom features like cap shape, color, gil size, odor etc. It is important to address data quality issues such as missing values and duplicates during preprocessing. The dataset should be divided into training & testing with relevant features selected for classification. The distribution of class labels should be considered, and any class imbalance issues should be handled appropriately. Data Transformers like Encoding techniques & PCA may be necessary to prepare data for the classification model.

2.7 Tools used

Python programming language and frameworks like NumPy & Pandas, Visualizing tools like matplotlib & seaborn, Machine Learning tools like scikit-learn, deployment tools like Flask along with interface design with HTML. VS code was used as IDE & Github as version control system.

- Python: A programming language that offers a vast ecosystem of libraries and frameworks making it suitable for various tasks.
- Numpy: Library for numerical computing
- Pandas: A powerful data manipulation and analysis library.
- Matplotlib: A plotting library that enables the creation of various types of high-quality plots & visualizations.
- Seaborn: Built on top of matplotlib, providing a high-level interface for creating attractive statistical graphics.
- Scikit-learn: A machine learning library, providing a comprehensive set of tools for various ML tasks like classification, regression etc.
- Flask: Lightweight & flexible web framework to build web application.
- HTML : HTML for creating web pages for describing the presentation and visual appearance of a webpage.
- Github: A web platform for version control and collaboration allowing user to store, manage & share their code repositories.
- VS code: A source code editor by Microsoft, providing versatile & customizable environment for coding in programming languages.

2.8 Constraints

An app must be user-friendly, with as automated as possible with users not being required to know any of the workings.

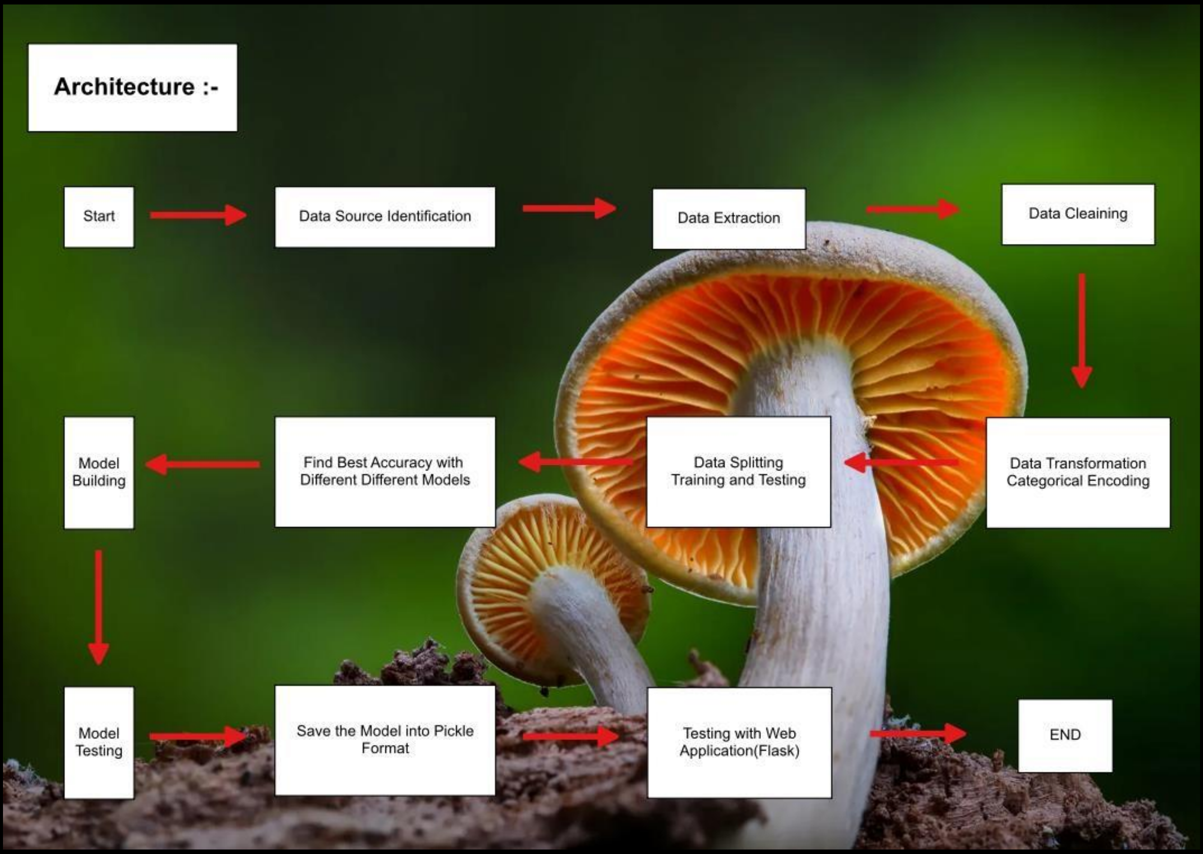
2.9 Assumptions

The main objective of the project is to implement classification based on the features of mushrooms. An ML-based classifier model is used for predicting the above-mentioned case. It is assumed that the user is able to identify the category of feature in mushroom.

3. Design Details

3.1 Process Flow

For classifying mushroom types, we use a classifier-based ML model. Below is process how it is work.


Architecture :-
Start
Data Source Identification
Data Extraction
Data Cleaining
Data Transformation Categorical Encoding
Data Splitting Training and Testing
Find Best Accuracy with Different Different Models
Model Building
Model Testing
Save the Model into Pickle Format
Testing with Web Application(Flask)
END

3.2 Event Log

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

Initial Step-by-step

- The System identifies at what step logging required
- The System should be able to log each and every system flow.
- Developer can choose logging method.
- System should not hang even after using so many loggings

3.3 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 Mushroom classifier solution is used for determining type of mushroom (edible or poisonous) based on inputting features. Mushrooms are use in edible purpose. In wide varieties of mushroom, it becomes necessary to know the mushroom edibility for the consumption basis. This program performs well giving almost 99% accuracy on the test set.

4.1 Reusability

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

4.2 Application Compatibility

The interaction with the application is done through the designed user interface, which the end user can access through any web browser

4.3 Resource Utilization

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

4.4 Deployment

Deployment can be done through local IP, also with cloud services like AWS, Azure or GCP.

Conclusion

The Designed Mushroom Classifier program will classify mushroom as edible or poisonous based on features of mushroom, which can help to identify mushroom's eat ability. This can prevent consuming poisonous mushrooms.

Thankyou