HIGH LEVEL DESIGN DOCUMENT(HLD)

MUSHROOM CLASSIFICATION

Revision No – 1.6

Last Date of Revision – 28/01/2023

Author : Sudhiram Chauhan

Document Version Control:

Date Issued	Version	Description
24/01/2023	1	Basic Formatting
24/01/2023	1.1	General Description
25/01/2023	1.3	Design Details
26/01/2023	1.4	Performance
27/01/2023	1.5	Conclusion
28/01/2023	1.6	Reference

CONTENTS

Abstract

Mushrooms have been consumed since earliest history. The word Mushroom is derived from the French word for Fungi and Mold. Now-a-days, Mushroom are popular valuable food because they are low in calories, carbohydrate, Fat, sodium and also cholesterol free. Besides this, Mushroom provides important nutrients, including selenium, potassium, riboflavin, niacin, Vitamin D, proteins and fiber. All together with a long history as food source. Mushroom are important for their healing capacity and properties in traditional medicine. It has reported beneficial effects for health and treatment of some disease. Many nutraceutical properties are described in Mushroom like cancer and antitumor attributes. Mushroom act as antibacterial, immune system enhancer and cholesterol lowering Agent. Additionally, they are important source of bio-active compounds. This work is a machine learning model that classifies mushrooms into 2 classes: Poisonous and Edible depending on the features of the mushroom. During this machine learning implementation, we are going to see which features are important to predict whether a mushroom is poisonous or edible.

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 before 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 Perspective

Our project Mushroom Classification is a Machine Learning based model that classifies mushrooms into 2 classes: Poisonous and Edible.

2.2   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.
The main goal is to predict which mushroom is poisonous & which is edible.

2.3   Proposed Solution

To solve the problem, we have created a User Interface for selecting the input from the user to predict whether the mushroom is poisonous or edible using our trained ML model. After processing the input, the last output (predicted value) from the model is communicated to the User.

2.4   Further Improvements

We analyzed the given data and extracted features that are important in predicting whether a mushroom is poisonous or edible. We can consider all other available features as well. However, it will result in a slower response on the web app that we have created. Also, since some of the characteristics of mushroom are same for poisonous and edible mushrooms, we recommend that User should also take help from someone who is expert in identifying edible mushrooms.

2.5   Technical Requirements

For technical requirements, we don't need any specialized hardware for virtualization of the application. The user should have a device that has the access to the web and the fundamental understanding of providing the input. And for the backend, we need a server to run all the required packages to process the input and predict desired output.

2.6   Data Requirements

A public dataset from Kaggle is used for our analysis purpose. This dataset was originally donated to the UCI Machine Learning repository. This dataset includes descriptions of hypothetical samples corresponding to 23 species of gilled mushrooms in the Agaricus and Lepiota Family Mushroom.

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

2.7   Tools Used

Python 3.9 is used as a programming language, Jupyter Notebook as IDE, Numpy, Pandas, seaborn, matplotlib, scikit learn for data preprocessing, data visualization and model building, HTML, CSS, Flask, Heroku for creating app and deploying model.


2.8   Constraints

The Mushroom Classification prediction answer should be user friendly, as automatic as attainable and also the user should not be needed to understand any of the operations.

3. Design Details

For identifying different types of anomalies in our data and for data preprocessing, we will use a machine learning base model. Below are the different process diagrams explaining the various steps that are involved in complete execution of this project.

3.1 Model Training & Evaluation

Mushroom Classification Architecture
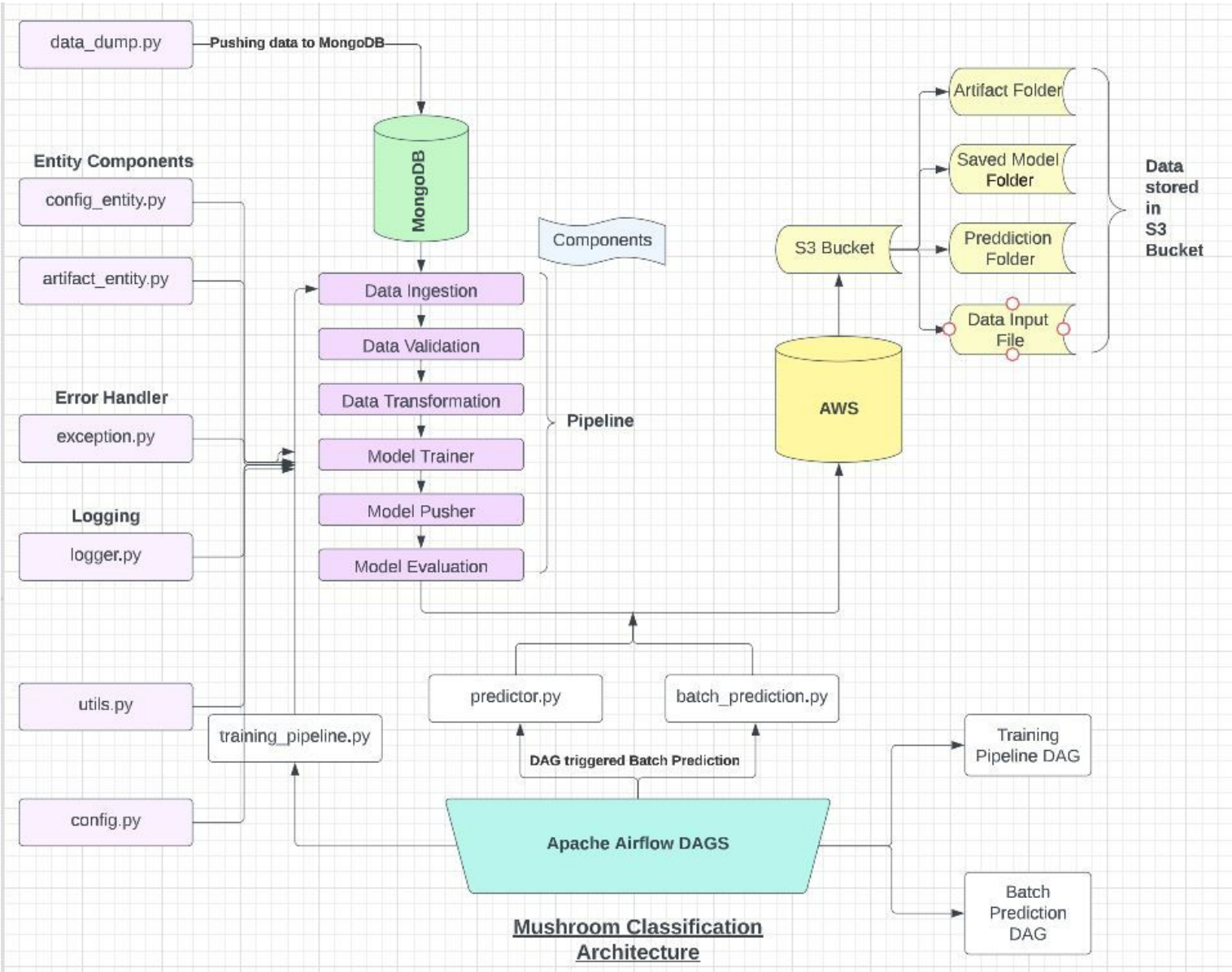
3.3   Logging

In logging, each time an error or an exception occurs, the event is logged into the system log file with reason and timestamp. This helps the developer to debug the system bugs and rectify the error.

3.4   Error Handling

Once the error occurs, the reason is logged into the log file with timestamp to rectify and handle it.

4. Performance

The Mushroom Classification solution is used for detecting whether a mushroom is poisonous or not, so it should be as accurate as possible. It is advised that a person should also take help from someone who is expert in identifying edible mushrooms since some of the characteristics of mushroom are same for poisonous and edible mushrooms.

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 different parts of the system are communicating or using Python as an interface between them. All the components have its own tasks to perform and it is the job of a Python to ensure proper transfer of data.

4.3   Resource Utilization

When a task is performed, it'll doubtless use all the process power offered till the process is finished.

4.4   Deployment

The model can be deployed using any cloud services such as Microsoft Azure, Amazon web services, Heroku, Google cloud, etc.

Conclusion

The designed Mushroom Classification system will detect whether a mushroom is poisonous or not based on various features that are present in the data. Hence, we can easily identify whether a mushroom is poisonous or not and will certainly help people in selecting right type of mushroom to consume.

References

1. INTERNATIONAL JOURNAL OF CLINICAL AND BIOMEDICAL RESEARCHEDIBLE MUSHROOM - A NUTRITIOUS FOOD IMPROVING HUMAN HEALTH - https://core.ac.uk/download/pdf/327998748.pdf

2. Wikipedia - https://en.wikipedia.org/wiki/Mushroom.