High Level Design

Project : Mushroom Classification

By : Sarthak Dehadray (FSDS 2022)

Version : 1.0 , Date :13/10/2022

**Document Version Control**

|  |  |  |  |
| --- | --- | --- | --- |
| **Person Name** | **Date** | **Version** | **Actions** |
| Sarthak Dehadray | 13/10/2022 | 1.0 | Document Created |
|  |  |  |  |
|  |  |  |  |

**Contents**

Document Version Control

Abstract

1. **Introduction**
   1. Why this HLD Document ?
   2. Scope
   3. Definitions
2. **General Description**
   1. Product Perspective
   2. Problem Statement
   3. Proposed Solution
   4. Further Improvements
   5. Data Requirements
   6. Tools Used
   7. Constraints
3. **Design Details**
   1. Process Flow
      * Model Training and Evaluation
   2. Event Log
   3. Error Handling
   4. Performance
   5. Reusability
   6. Application Compatibility
   7. Resource Utilization
   8. Deployment
4. **Conclusion**

**Abstract**

In the current scenario the consumption of Mushrooms is on the rise and thus is the cultivation of mushrooms as well. The farmers at the time of harvest are posed with an unprecedented problem of **“Identifying Edible or Non-Edible Mushrooms”**

**About the Dataset**

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 and this work discusses the utility of Machine Learning in solving this problem.**

**1) Introduction**

**Why this High Level Design Document?**

This document will showcase the details of the current project ,and, also represent the suitable model for coding. This document can be used to streamline the flow of the provided solution and its viability to solve the given problem. This document also serves as a reference manual for understanding the implementation at a high level.

**This HLD will :**

* 1. Present all of the design aspects and define them in detail.
  2. Describe the UI.
  3. Describe Hardware and Software Interfaces.
  4. Describing the Performance Requirements.
  5. Include the design features and architectures of the project.
  6. List of other utilities like :

Security

Reliability

Maintainability

Portability

Reusability

Application Compatibility

Resource Utilization

Serviceability

**Scope**

The HLD tries to present the structure of system such as database architecture, application architecture, application flow , and technology architecture. This HLD is written in layman terms.

**Definitions**

* + Database : Collection of Information/ Data.
  + Dataset : The information about Mushrooms along with features in a tabular format.
  + IDE : Integrated Development Environment.

1. **General Description**

**Perspective**

The model will help in detecting the edibility of mushroom in an automated way i.e without any human intervention.

**Problem Statement**

To create a ML solution and automate decision making process for:

* + - Preparing a model based on given features to classify edible or non edible mushrooms.
    - This model will formulate rules for segregation of mushrooms.

**Proposed Solution**

The solution proposed is a **Classification Model** built using different Machine Learning Algorithms which are suitable for the dataset.

**Further Improvements**

These models, once trained on sufficient amount of data can be used to integrate this process of segregation in real time and on a production line.

**Data Requirements**

The training will need :

* + - Data having all features necessary for classifying a sample as Edible or Non Edible.
    - The dataset should have labels as it is required for training the Machine Learning model.
    - Following formats are acceptable
      * Excel File
      * CSV File
      * Json file

**Tools Used**

* + - Python
    - Pandas
    - Mongo DB
    - Numpy
    - SciKit Learn
    - Flask
    - Python Classic IDE
    - Heroku / Azure for deployment
    - Github

**Constraints**

The model should not classify a non edible mushroom as an edible one, also , the model should work on various species of mushrooms.

1. **Design Details**

**Process Flow**

Data Ingestion

Training / Validation

To ML Model

Prediction Edible/ Not

New Data

**Model Training Evaluation**

Training Data

Prediction

Model

Test Data

Augmented Data

Result/ Output

Dataset

**Event Log**

The system should be able to log each and every relevant information.

Necessary logging files should be created.

Logging is needed for smooth debugging process.

**Error Handling**

Any misbehavior in the flow should be handled. Error might occur due to invalid data , user interruption or any other event . Logging is necessary for any such event.

**Performance**

The model should accurately classify non edible mushrooms.

**Reusability**

Modular coding is necessary for implementation of the project.

**Compatibility**

The application should be compatible with all the necessary user interfaces.

**Resource Utilization**

The process will use all the processing power for completing the task.

**Deployment**

Heroku / Microsoft Azure

**Conclusion**

**The app will be able to detect Non Edible Mushrooms in an automated manner and would also formulate the rules for taking the decision.**