

Mushroom Classification

High Level Design Document

Revision Number: 2.0

Last date of revision: 12/04/2023

Contents

Abstract.....	4
1. Introduction	5
1.1 Why this High Level Design Document	5
1.2 Scope	5
2. General Description	6
2.1 Product Perspective	6
2.2 Problem Statement	6
2.3 Proposed Solution	6
2.4 Technical Requirements.....	6
2.5 Data Requirements	6
2.6 Tools Used.....	6
2.6.1 Hardware Requirements	7
2.7 Constraints	7
3. Design Details.....	8
3.1 Process Flow.....	8
3.1.1 Model Training and Evaluation.....	8
3.1.2 Deployment Process.....	9
4. Performance	10
4.1 Reusability.....	10
4.2 Application Compatibility.....	10
4.3 Resource Utilization	10
4.4 Deployment.....	10

Document Version Control

Date Issued	Version	Description	Author
19/03/2023	1.0	Initial HLD – V1.0	Nupur Mishra
12/04/2023	2.0	Updated	Nupur Mishra

Abstract

Mushrooms are widely known for their great taste and amazing health benefits. Packed with a ton of essential vitamins and minerals, they make for an excellent addition to your diet, adding flavor to many different recipes.

Mushrooms are a low-calorie food that packs a nutritional punch. Loaded with many health-boosting vitamins, minerals, and antioxidants, they've long been recognized as an important part of any diet.

In a lot of countries, mushrooms are abundant and people collect wild species for their unique taste and health benefits. Mushroom hunting(also known as shrooming) is a free activity for people who like exploring but not all of them are edible. Some species are poisonous as they contain different types of naturally occurring toxicants, and it can be risky even for people skilled with identifying them.

Hence, we try to identify which mushrooms are edible by creating a classification model and using the dataset(Mushroom Classification) that was originally contributed to the UCI Machine Learning repository

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:
 - Security
 - Reliability
 - Maintainability
 - Portability
 - Reusability
 - Application compatibility
 - Resource utilization
 - Serviceability

1.2 Scope

The HLD documentation presents the structure of the system, such as 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 Product Perspective

This application is useful to classify whether the mushroom is edible or poisonous, which will help to prevent humans from dangerous

2.2 Problem Statement

To create a machine learning model to classify if the mushroom is edible or not.

2.3 Proposed Solution

A machine learning model is to be implemented to perform the mushroom classification. It will train the model and get the input from UI it is required to predict the category of the mushroom, which is supposed to be classified. The input will be processed and given to the model to predict the output, then the output will be displayed to the web page, where they started the request.

2.4 Technical Requirements

This application is based on python language, so python should be installed and it should be properly configured with the PATH.

- The application should have the access to use the python environment
- Necessary packages listed in the requirements.txt should be made available for this application

2.5 Data Requirements

This application uses very less data when compared to other applications which are in the market now. It only carries 22 categorical data, which only requires very less data.

2.6 Tools Used

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

- VSCode is used as Code Editor
- Jupyter notebook is used for data analysis
- For visualization of the plots Matplotlib and Seaborn were used.
- Front end development is done using HTML/CSS
- Python Flask library is used for backend development.

- GitHub is used as a version control system.
- AWS Beanstalk is used for deployment of the model.



2.6.1 Hardware Requirements

PC or Laptop or any remote server which can run the application with necessary software which are mentioned in the technical requirements.

2.7 Constraints

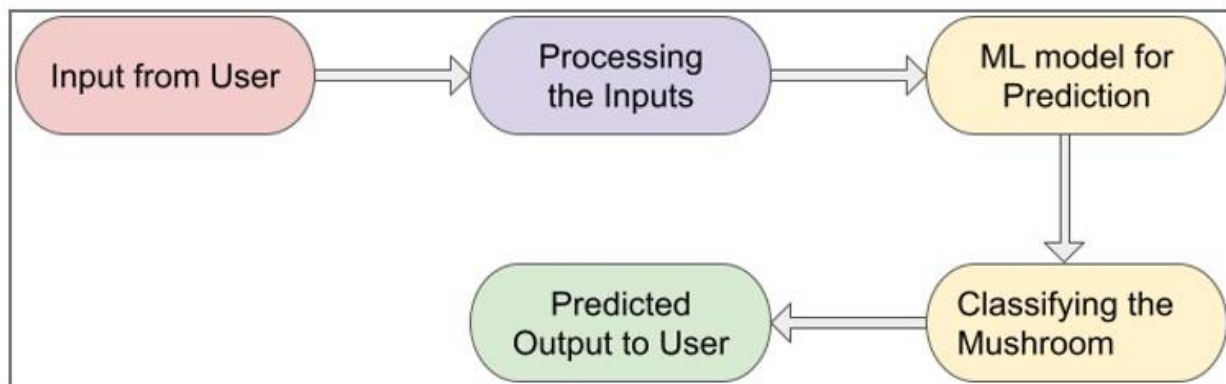
This classification system must be user friendly, and the user should not be required to know any of the workings of ML models.

3. Design Details

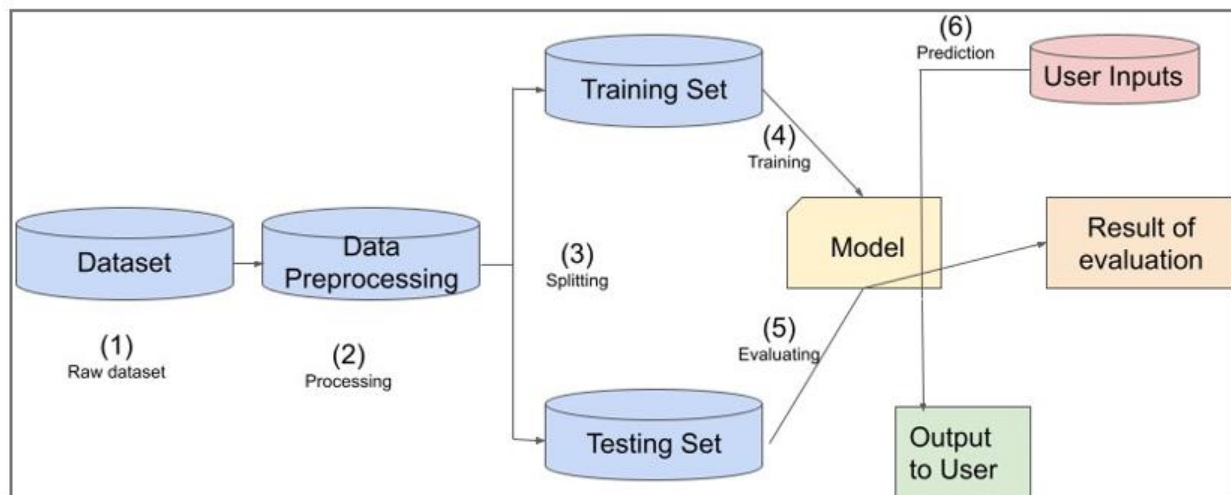
3.1 Process Flow

For identifying the different types of anomalies, we will use a deep learning base model. Below is the process flow diagram as shown below.

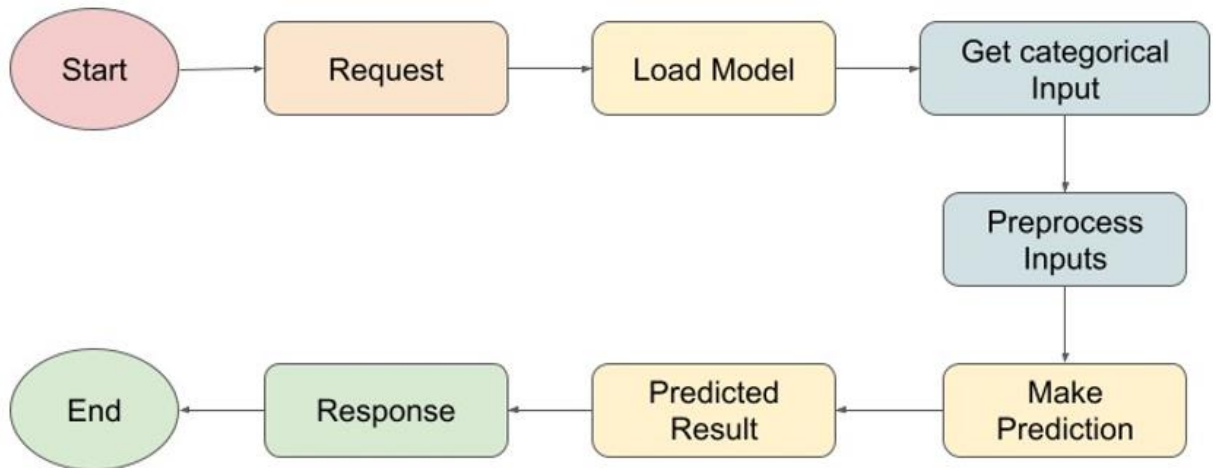
Proposed methodology



3.1.1 Model Training and Evaluation



3.1.2 Deployment Process



3.2 Event log

The system should identify where logging is required. It should maintain logs for different steps of code for reference during program debugging. Logging is to be maintained in files.

System should not hang even after using too many loggings.

3.3 Error Handling

Exceptions (Anything that happens outside the normal and intended usage) should be handled for every part of code. They should be recorded in log file for reference.

4. Performance

This classifier is used for classifying whether the mushroom is edible or poisonous when providing the required inputs. The model should be performing well with good accuracy and recall.

4.1 Reusability

The code written to check different models based on training is in a modular form like Object. So, it can be reused wherever and whenever needed.

4.2 Application Compatibility

All the requirements are listed in the requirements.txt for packages. And the software required for running the packages is Python 3.X >= Python 3.6

4.3 Resource Utilization

When a request is made, then it will utilize only the main thread. So, it will utilize less resources.

4.4 Deployment

This project is to be deployed on AWS Beanstalk using Code pipeline.



Conclusion

The model should be able to identify if a mushroom is edible or not based on its features. The model should be able to accept input from users and data should be validated.

Also, the model should have minimum 'Recall' as classifying poisonous mushrooms as edible can be dangerous.

This will be helpful for users to identify which mushrooms should they consume.