HIGH LEVEL DESIGN DOCUMENT

MUSHROOM CLASSIFICATION

AAKASH BHUTE

CONTENTS

Abstract

This study focuses on the classification of mushrooms into two categories, namely Poisonous and Edible, using a machine learning model. It aims to determine the significant features that play a crucial role in predicting the edibility or toxicity of mushrooms. Mushrooms have been consumed since ancient times and are highly regarded for their nutritional value. They are low in calories, carbohydrates, fats, and sodium, while being cholesterol-free. Mushrooms offer essential nutrients such as selenium, potassium, riboflavin, niacin, Vitamin D, proteins, and fiber. They have a rich history as a food source and are also recognized for their healing properties in traditional medicine. Various health benefits and potential disease treatments have been associated with mushrooms, including their anticancer and antitumor properties. Moreover, mushrooms exhibit antibacterial effects, enhance the immune system, and assist in lowering cholesterol levels. Furthermore, mushrooms are a valuable source of bioactive compounds. Throughout this machine learning analysis, we will identify the key features that determine whether a mushroom is poisonous or edible.

1. Introduction

1.1 What is a 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, utilizes a machine learning model to categorize mushrooms into two distinct classes: Poisonous and Edible.

2.2 Problem Statement

The main objective of our project, Mushroom Classification, is to accurately predict whether a mushroom is poisonous or edible. To achieve this, we utilize the information provided in the Audubon Society Field Guide to North American Mushrooms. The guide includes descriptions of hypothetical samples for 23 species of gilled mushrooms from the Agaricus and Lepiota families, categorizing them as definitely edible, definitely poisonous, or potentially edible but not recommended (which we merge with the toxic category). It is important to note that the guide emphasizes the absence of a simple rule to determine a mushroom's edibility, unlike the "leaflets three, leave it be" guideline for Poisonous Oak and Ivy.

2.3 Proposed Solution

In order to address this problem, we have developed a User Interface that allows users to input mushroom details for classification as either poisonous or edible using our trained machine learning model. The user-provided information is processed, and the final output, which represents the predicted classification, is communicated back to the user.

2.4 Further Improvements

After conducting an analysis of the provided data, we identified and extracted the significant features that contribute to predicting whether a mushroom is poisonous or edible. While it is possible to consider all available features, incorporating them might result in slower response times

on the web application we have developed. It is important to note that certain characteristics of mushrooms can be shared between both poisonous and edible varieties, so we strongly advise users to seek assistance from an expert in mushroom identification to ensure accurate categorization.

2.5 Technical Requirements

No specialized hardware is necessary for virtualizing the application. Users only need a device with internet access and a basic understanding of how to provide input. On the backend, a server is required to execute the necessary packages for processing the input and generating the desired output.

2.6 Data Requirements

For our analysis, we utilized a publicly available dataset sourced from Kaggle. The original source of this dataset is the UCI Machine Learning repository. It contains detailed descriptions of hypothetical samples, specifically pertaining to 23 species of gilled mushrooms within the Agaricus and Lepiota Family Mushroom categories.
Dataset :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.

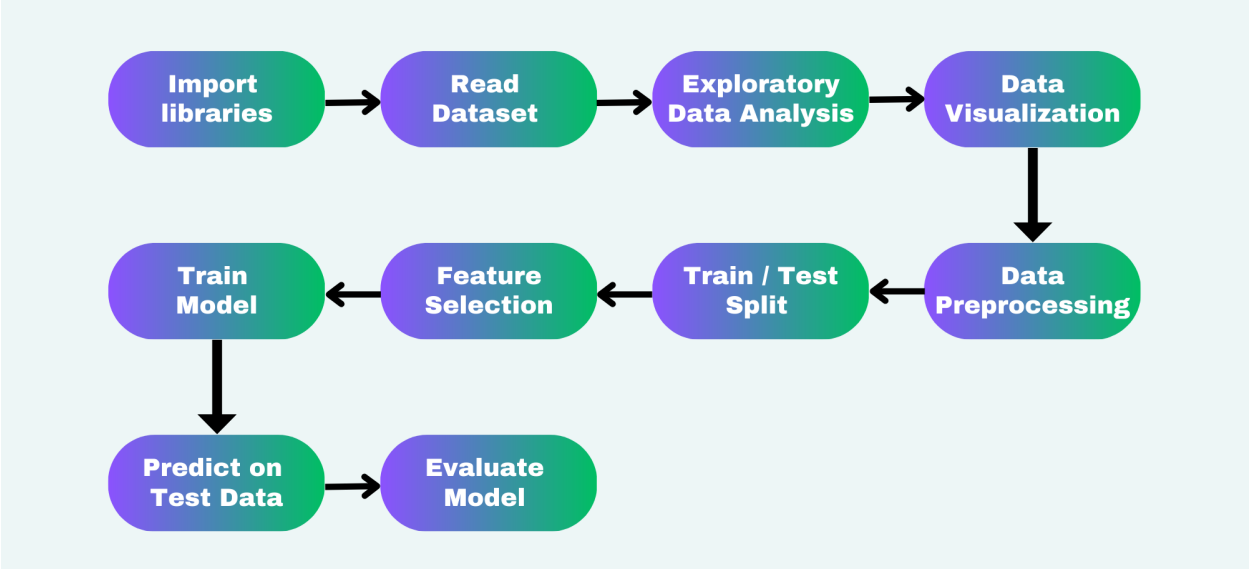
2.8 Constraints

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.

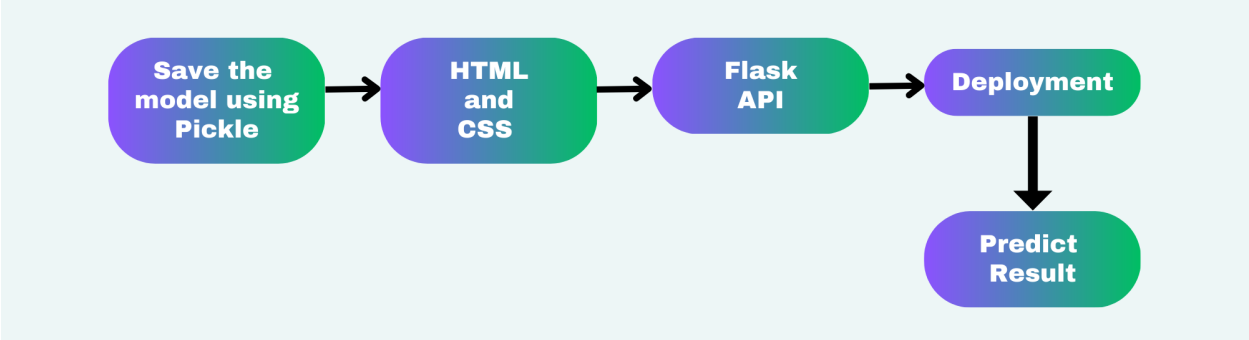
3. Design Details

To identify various types of anomalies in the data and perform data preprocessing, a machine learning-based model will be employed. The project encompasses multiple process diagrams that illustrate the various steps involved in the complete execution of this project.

3.1 Model Training & Evaluation



3.2 Deployment Process



3.3 Logging

In the logging process, whenever an error or exception occurs, the event is recorded in the system log file along with its reason and timestamp. This logging mechanism facilitates developers in debugging system bugs and resolving errors by providing valuable information about the occurrences.

3.4 Error Handling

Upon the occurrence of an error, the reason for the error is logged into the log file along with a timestamp. This allows for proper identification and handling of the error, aiding in the process of rectifying any issues that may arise..

4. Performance

The Mushroom Classification solution is designed to accurately detect whether a mushroom is poisonous or edible. However, due to the presence of certain shared characteristics between poisonous and edible mushrooms, it is strongly advised that individuals seek assistance from an expert in mushroom identification. This additional guidance will help ensure the highest level of accuracy in determining the edibility of mushrooms.

4.1 Reusability

The code and components implemented in the solution are designed to be reusable without encountering any issues. They are built with modularity in mind, allowing for easy reuse in future projects or extensions of the current solution.

4.2 Application Compatibility

The various components of the system communicate and interact with each other through Python, which serves as the interface between them. Each component has its own defined tasks and responsibilities, and it is the role of Python to facilitate the seamless transfer of data between these components, ensuring efficient communication and collaboration within the system.

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 Mushroom Classification system is specifically designed to determine whether a mushroom is poisonous or not by leveraging multiple features present in the data. This enables us to accurately identify the edibility of mushrooms and significantly aids individuals in making informed decisions when selecting mushrooms for consumption.