**Mushroom Classification**

**based on Visual Description**

Revision Date: 18/07/2024

|  |  |  |  |
| --- | --- | --- | --- |
| **Date Issued** | **Version** | **Description** | **Author** |
| 16/07/2024 | 1.0 | Initial Project | Arush James |
| 18/07/2024 | 1.1 | Updated UI | Arush James |
|  |  |  |  |

Contents:

1: Introduction

1.1: Brief Introduction

1.2: Definitions

2: General Description

2.1: Problem Statement

2.2: Proposed Solution

2.3: Improvements on Initial Solution

2.4: Technical Requirements

2.5: Data Requirements

2.6: Tools Used

2.7: Assumptions

3: Design Details

3.1: Process Flow

3.2: Reusability

3.3: Resource Utilisation

3.4: Deployment

4: Conclusion

4.1: Conclusion for Project

**1)Introduction**

1.1: Brief Introduction:

A mushroom or toadstool is fleshy, [spore](https://en.wikipedia.org/wiki/Spore)-bearing [fruiting body](https://en.wikipedia.org/wiki/Sporocarp_(fungi)) of a [fungus](https://en.wikipedia.org/wiki/Fungus), typically produced above ground, on soil, or on its [food](https://en.wikipedia.org/wiki/Food) source. Toadstool generally denotes one poisonous to humans. Mushrooms are also consumed a lot by humans due to its “umami” or “shitake” like taste. So therefore, it is incredibly important to distinguish poisonous mushrooms from non-poisonous ones as consuming the latter might cause unwanted symptoms or in a worst-case scenario, spell death. But if we able to find out the different characteristics of the mushroom, it will be possible to find out which ones are non-toxic

1.2: Definitions:

-Logistic Regression: An algorithm used to predict categorical data.

-Modules: It is a reusable block of code in a python file containing functions, classes, and other definitions.

-Accuracy: Accuracy in machine learning is simply the percentage of correct predictions a model makes.

-Test Data: It is unseen data used to evaluate a model's general performance on new information.

-Training Data: It's the information the model learns from to make predictions.

-Package: A package is a directory that groups related Python modules (reusable code files) together

**2)** **General Description**

2.1: Problem Statement:

As discussed above in chapter 1.1, mushrooms are often consumed by humans. The problem arises when we are not able to differentiate between poisonous and non-poisonous ones.

The goal of this project is to develop a Mushroom classifier with a User-Friendly UI that helps us to predict if a mushroom is poisonous or not using Python and Machine Learning.

2.2: Proposed Solution:

In order to develop such an application, we would need a dataset. This can easily be found in the publicly available, UCI Machine Learning Repository dataset containing mushroom descriptions and classifications. The dataset used in this project includes descriptions of hypothetical samples corresponding to 23 species of gilled mushrooms in the Agaricus and Lepiota Family Mushroom drawn from The Audubon Society Field Guide to North American Mushrooms (1981). Each species is identified as definitely edible, definitely poisonous, or of unknown edibility and not recommended. This latter class was combined with the poisonous one.

Furthermore, we would need to choose a Machine Learning (ML) Algorithm suitable for binary classification. The ideal ML model for such a classification would be Logistic Regression.

To train the model, we used 80% of the dataset. For testing, we used 20% of the dataset. After testing, we got an accuracy of almost 99.8%.

For the UI, we settled on Flask. Flask was chosen because:

-Flask offers more control over the application's structure and functionality. Especially since we can easily implement more HTML, CSS and JavaScript functionality to the application.

-Flask can easily be scaled up.

The main problem encountered when using Flask was:

-It required in depth knowledge of Web-Development like Routing and Templating.

-The issue of getting ambiguous errors.

2.3: Improvements on Initial Solution

During Development, it was found that having a homepage for the application was a necessity. This was because the homepage would guide the user on how to use the application to its maximum potential.

This feature would boost user-friendliness.

2.4: Technical Requirements

* **Hardware (Recommended)**

1. **Processor:** Any mid-range multi-core processor
2. **RAM:** 6GB RAM
3. **Storage:** Above 128GB

* **Software (Recommended)**

1. **OS:** Anything above Windows 7
2. **Python:** Any python version after Python 3.8
3. **Browser:** Any Web browser is fine

2.5: Data Requirements

The dataset used was a csv file from UCI Machine Learning Repository dataset.

Link: <https://www.kaggle.com/datasets/uciml/mushroom-classification>

2.6: Tools Requirements

A few python tools(modules) for different purposes:

* **Data Cleaning and Preprocessing**

1. Pandas (Version: 2.2.1)
2. NumPy (Version: 1.26.4)

* **EDA (Exploratory Data Analysis)**

1. SNS Seaborn (Version: 0.13.2)
2. Matplotlib (Version: 3.9.1)

* **User Interface**

1. Flask (Version: 3.0.3)
2. HTML5
3. Bootstrap 5

* **Machine Learning**

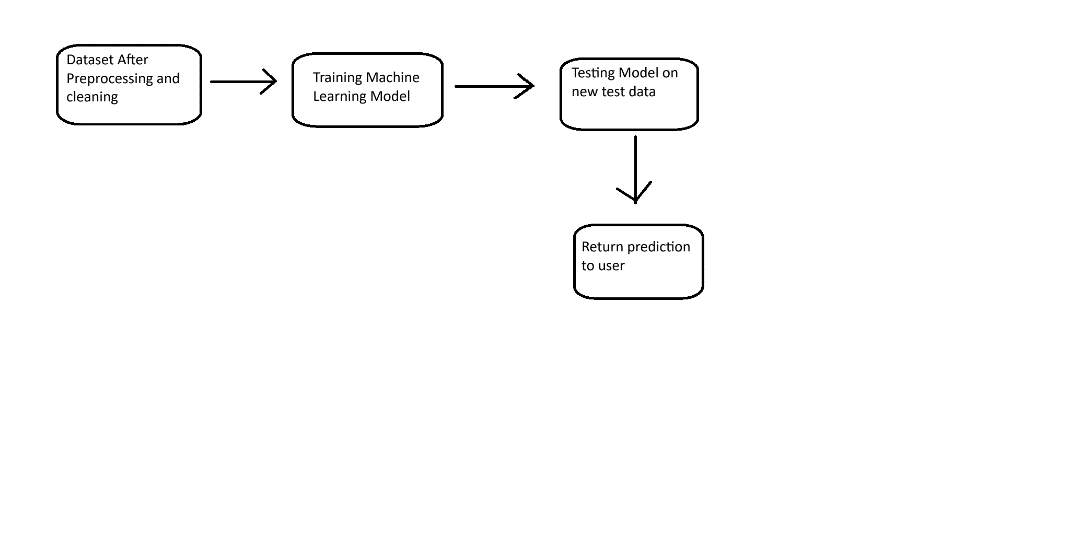
1. Scikit-Learn (Version: 1.5.1)

2.7: Assumptions

We **assume the dataset's information is accurate and comprehensive,** representing the variety of mushrooms we aim to classify, and we also **assume the dataset encompasses a sufficient range of mushroom characteristics**, allowing the model to generalize well to new mushrooms.

**3)Design Details**

3.1: Process Flow

****

3.2: Reusability

The application after multiple tests still ran successfully with a high accuracy. So, the application can be reused multiple times.

3.3: Resource Utilisation

Training the machine learning model will require processing power (CPU/GPU) and memory (RAM).

3.4: Deployment

For deployment of the project, we can use AWS, Heroku and Azure.

**4) Conclusion**

4.1: Conclusion for Project

This project successfully developed a mushroom classifier application utilizing Logistic Regression model. The model achieved a 99.8% on the test data, demonstrating its potential for identifying different mushroom types. The user interface allows users to input various mushroom characteristics and receive classification results.