PROJECT REPORT

MUSHROOM CLASSIFICATION

By - Prabhjot Singh

CONTENTS

1. Abstract

Mushrooms hold a timeless significance in human gastronomy, intertwined with both mystery and familiarity. Their name originates from French, linking them to fungi and mold, carrying an air of enigma. Today, mushrooms are valued for their nutrition, low calorie content, and absence of cholesterol, making them popular for health-conscious eaters.

This project introduces an advanced Mushroom Classification Machine Learning Model capable of accurately categorizing mushroom species as either poisonous or edible. Through a meticulously curated dataset and advanced neural networks, the model identifies distinctive patterns for precise classification. Its user-friendly interface accommodates users of all expertise levels, while its adaptability ensures robust performance across various conditions. This innovation not only revolutionizes mushroom classification but also provides a vital tool for mycologists, researchers, and enthusiasts. The report outlines the model's architecture, development, and real-world applicability, highlighting its significant contribution to the field of mycology.

2. Introduction

The Detailed Project Report (DPR) serves as a comprehensive blueprint for project success, encompassing resources, tasks, and strategies. This report covers the end-to-end implementation of the Mushroom Classification project, including architecture, data visualization, preprocessing, model building, performance evaluation, and deployment. Sample test cases provide practical insights, bridging theory and application. The DPR is a valuable resource, encapsulating the project's journey from conception to realization.

3. Problem Statement

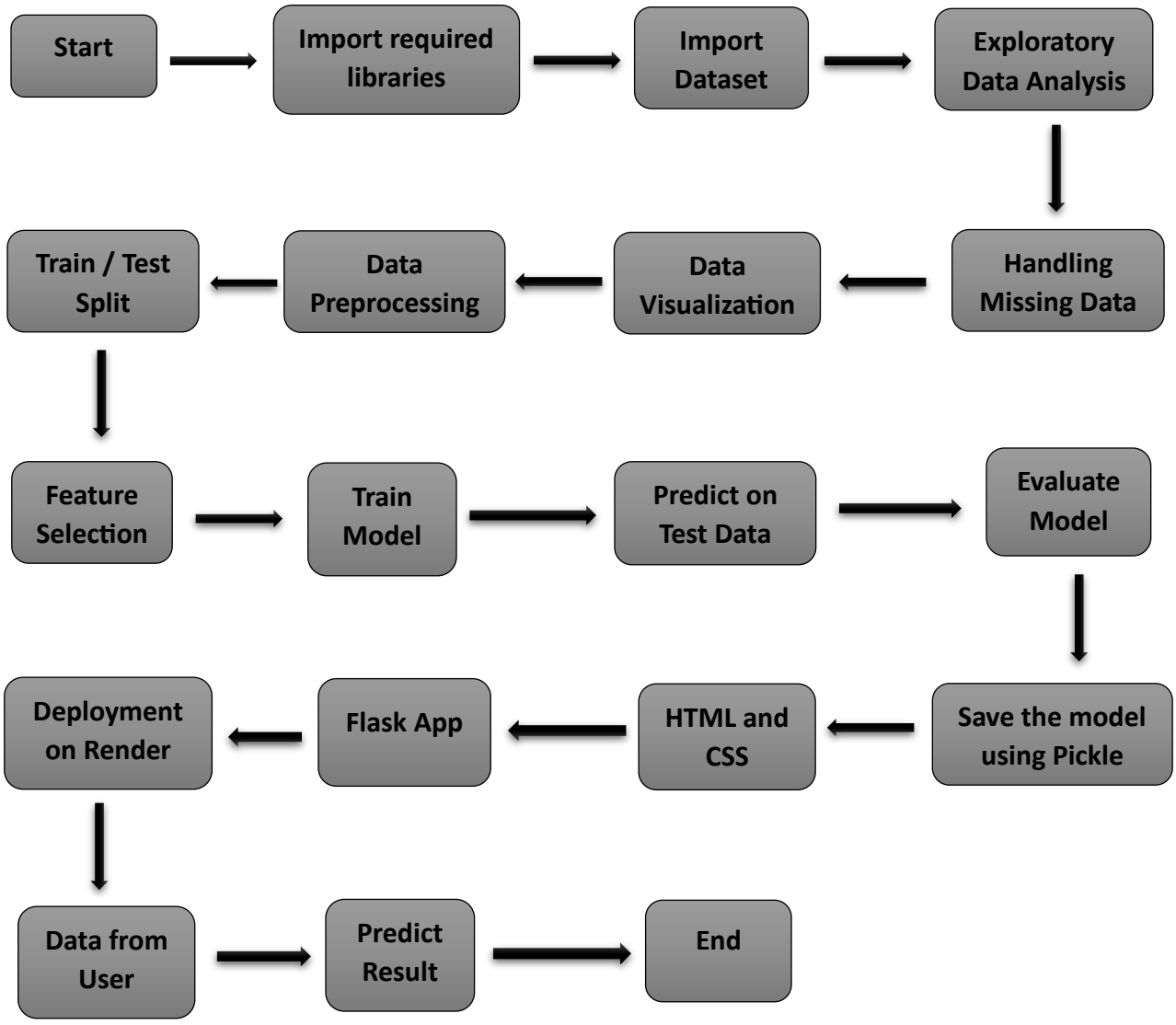
The Audubon Society Field Guide to North American Mushrooms features descriptions of hypothetical samples, encompassing 23 gilled mushroom species within the Agaricus and Lepiota family (1981). These species are categorized as definitely edible, definitely poisonous, or potentially edible but not recommended – the latter of which is combined with the toxic category. The guide emphasizes that no straightforward rule akin to "leaflets three, leave it be" for Poisonous Oak and Ivy exists to judge mushroom edibility.

The central objective is to predict whether a mushroom is poisonous or edible. This prediction task forms the crux of the project, involving the utilization of various data-driven techniques to decipher the complex patterns that determine a mushroom's classification. By leveraging machine learning and predictive modeling, the project aims to provide a tool that assists in making informed decisions about the safety of consuming mushrooms based on their attributes.

4. Tools

As we forge ahead, Python 3.9 assumes the mantle of our programming prowess. Jupyter Notebook becomes the canvas of ideation. The artistry of data preprocessing and visualization unfolds with NumPy, Pandas, Seaborn, and Matplotlib. Sklearn ushers in the dawn of model creation. In the realm of deployment, HTML, CSS, Flask, and Render converge to craft an interactive oasis.

5. Architecture



6. Architecture Description

At the core of this project lies a user-centric interface, ingeniously designed to empower users with the ability to discern the toxicity of mushrooms. This architecture orchestrates a symphony of data collection, exploration, preprocessing, modeling, and deployment, all harmonizing to craft a seamless and accurate prediction experience.

6.1 Data Procurement

The wellspring of data for this endeavor is drawn from a Kaggle Dataset, a repository renowned for its diversity and quality. The dataset, accessible through the following URL, encapsulates descriptions of hypothetical samples encompassing 23 distinct species of gilled mushrooms, derived from the esteemed "Audubon Society Field Guide to North American Mushrooms" (1981):

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

6.2 Data Insight

A symphony of mushroom taxonomy unfolds within this dataset, featuring 8124 records across 23 categorical attributes. A pivotal column emanates as the fulcrum of prediction—its contents resonating with the binary distinction of 'p' for poisonous and 'e' for edible. The data's equilibrium is palpable, with a nearly equivalent distribution of the two classes.

6.3 Illuminating Patterns

Navigating the data's intricacies, a canvas emerges through exploratory data analysis, exhibiting discernable trends and insights. To unveil the essence of the mushroom's classification, the focus narrows to selected features, a calculated subset from the ensemble. Visual narratives—count plots—effuse insights into the delicate balance of our classification target.

6.4 Filling the Gaps

Intricacies unravel further, as missing data demand attention. The keen eye identifies the 'stalk-root' column harboring these vacancies. Amidst this intricate tapestry, the question mark takes a semblance of imperfection. The remedy, an artful one—2490 instances of uncertainty yield to the crafted hand of the Simple Imputer, restoring coherence to this mosaic.

6.5 Visualizing Insights

Visualization, a bridge between numbers and understanding, becomes a quintessential endeavor. The chosen attributes resonate with the target column, each plot an incandescent glimpse into the mushroom's soul. Trends form, secrets emerge, and a deeper connection to the data is forged.

6.6 Crafting Preparedness

The journey through preprocessing begins with discernment—the 'veil-type' column, a monotonous echo, exits the stage. The target column transforms, the symbiotic embrace of Label Encoder yielding numerical reflection to categorical essence. Scaling emerges as the unifier, bringing disparate features to the harmonious cadence of a shared class.

6.7 The Quest for Relevance

With data primed, the feature selection phase illuminates—the SelectKBest method invokes chi-squared scrutiny, unraveling a constellation of 12 attributes that harmoniously resonate with our predictive goal. This ensemble, handpicked for its resonance with the target, shapes the essence of model training.

6.8 Orchestrators of Prophecy

Stepping forth as the quintessential arbiters of prescience, the Random Forest Classifier emerges as a beacon of predictive supremacy. From the dim recesses of data's labyrinth, it ascends with unparalleled swiftness. Mastery over the data's intricate dance culminates in an astonishing declaration—both training and test domains stand adorned with the resplendent mantle of 100% accuracy. The symphony of data science reaches its crescendo, the model's resounding harmonies resplendent in their assurance.

6.9 Illuminating the Virtual Realm

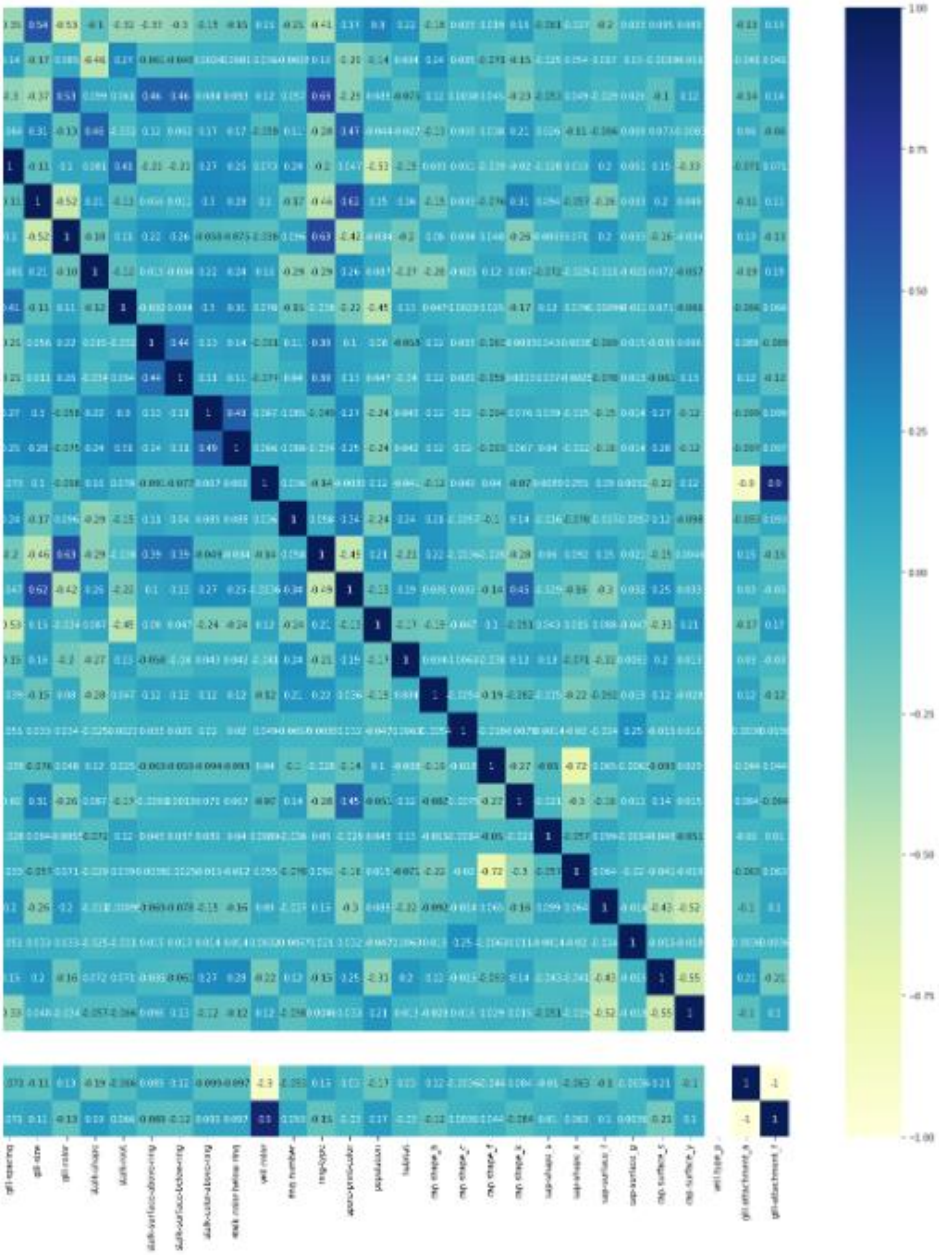
The apogee of our pursuit materializes, a virtual sanctuary fashioned through the deft artistry of HTML and CSS. This digital canvas, a realm unto itself, blossoms into the sanctum for our Flask web application. Like a newborn star, it embarks on its journey, from local incubation to the warm embrace of Render—a culmination that resounds with flawless deployment. Inputs tendered are met with the contemplative embrace of algorithms, yielding outcomes that mirror verity—a digital oracle meticulously bound with threads of precision.

In this intricate architecture, the harmonious choreography of data and technology entwines itself with the enigma of mushrooms' classification. A symphony of predictive finesse arises, its echoes resonating from the very core of data to the very touchpoints of users' engagement. This is an embrace of experience, where the nuanced interplay of insight and refinement coalesce, cocooning users in a realm that is as enlightening as it is elegant.

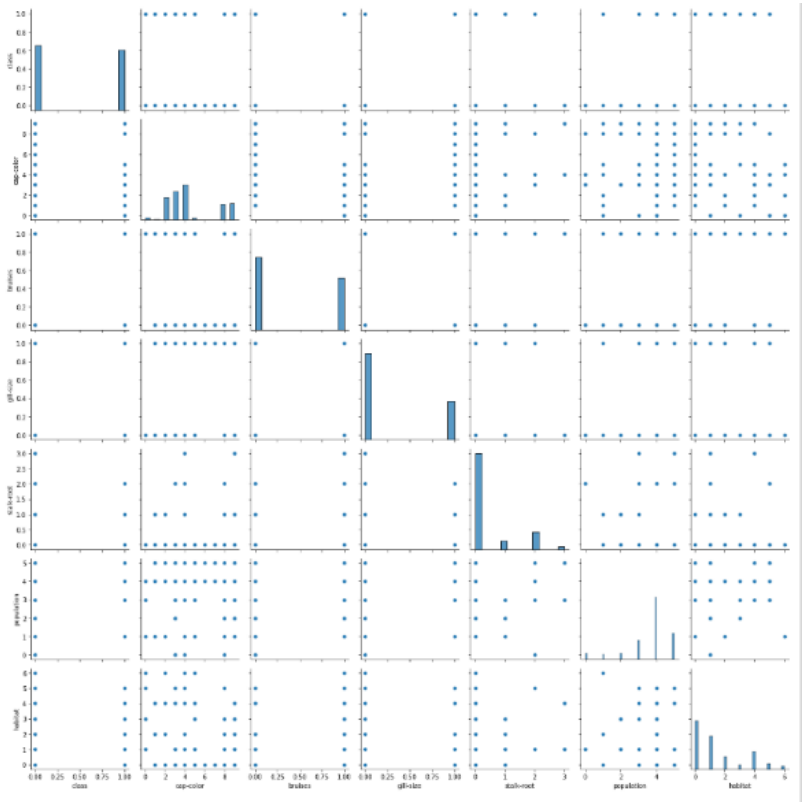
7. Data Visualization

7.1 Heatmap

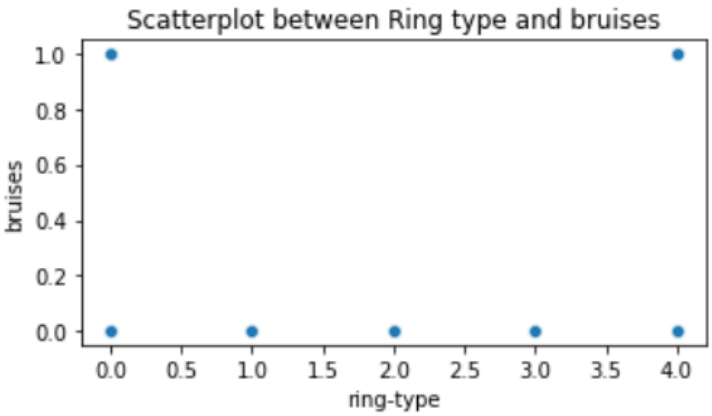
In the dataset, "gill attachment_f" and "gill attachment_a" show strong correlations with "veil color," with one positively and the other negatively correlated. Analyzing their correlation with the target "class," "veil color" emerges as more significant. Thus, "gill attachment_f" and "gill attachment_a" could be dropped. Furthermore, "veil type_P" holds a constant value across all rows, making it non-informative and suitable for removal, streamlining the dataset for analysis.

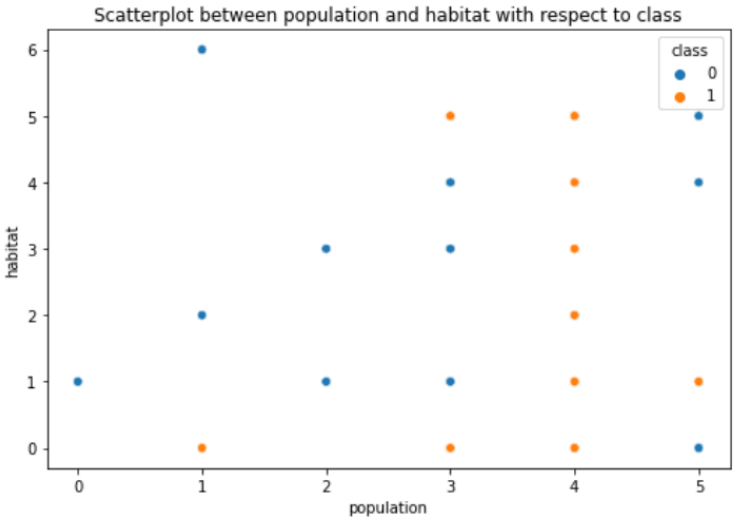


7.2 Pair plot

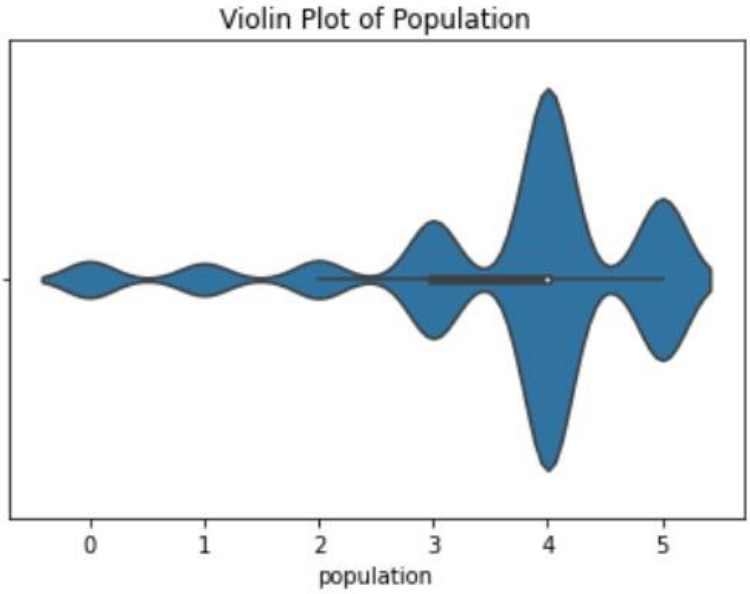


7.3 Scatter plot

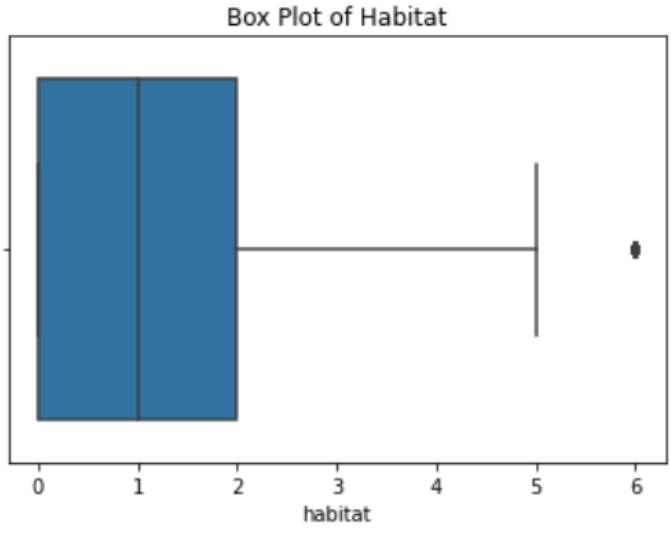


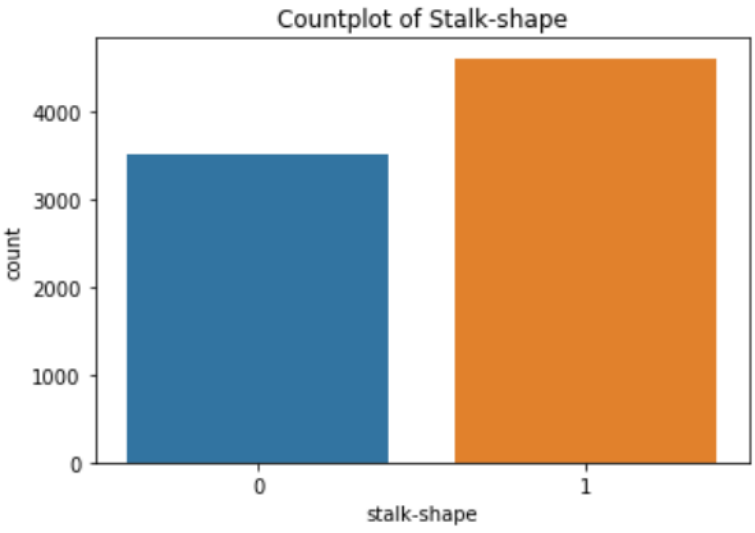


7.4 Violin plot

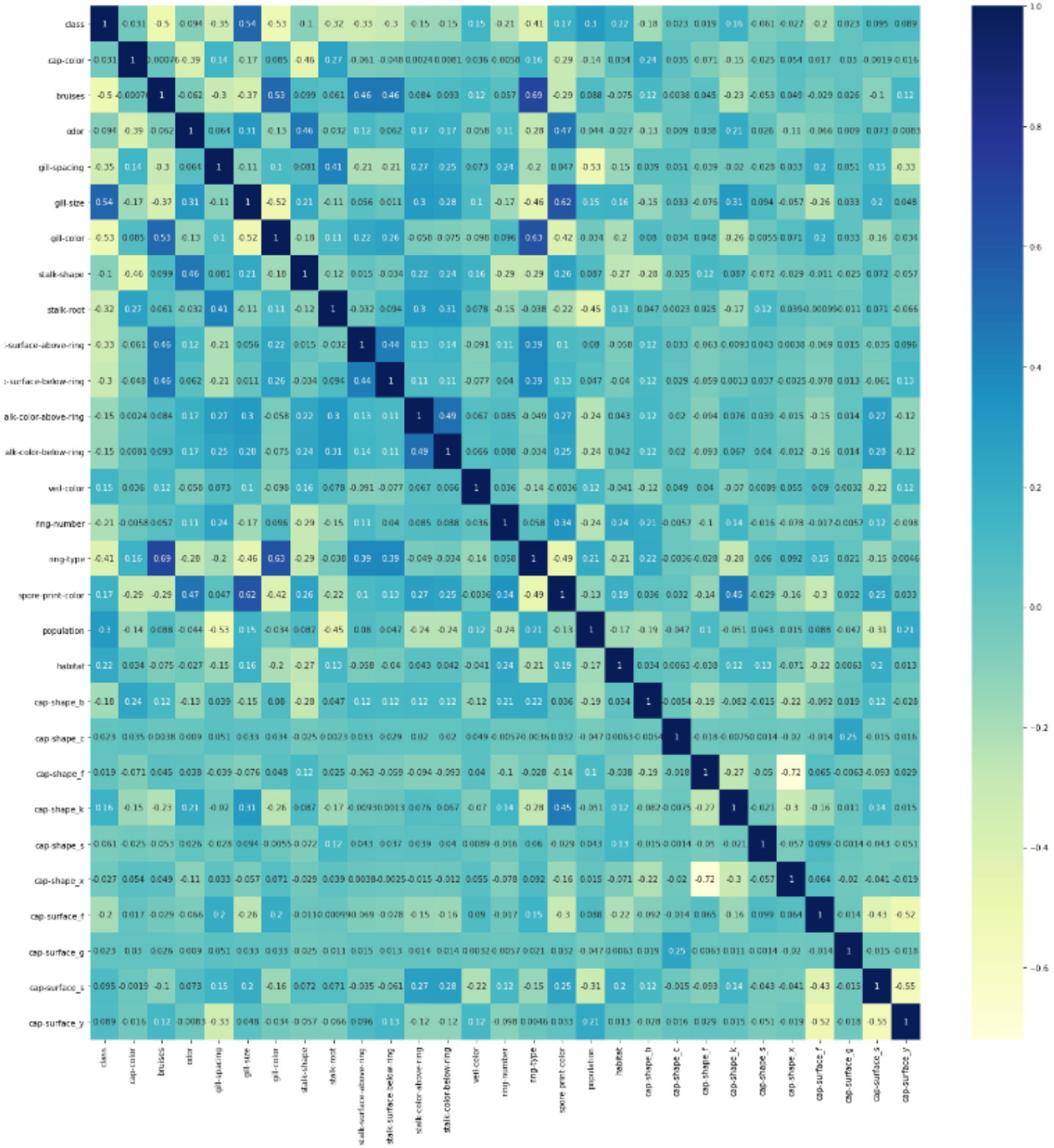


7.5 Box plot

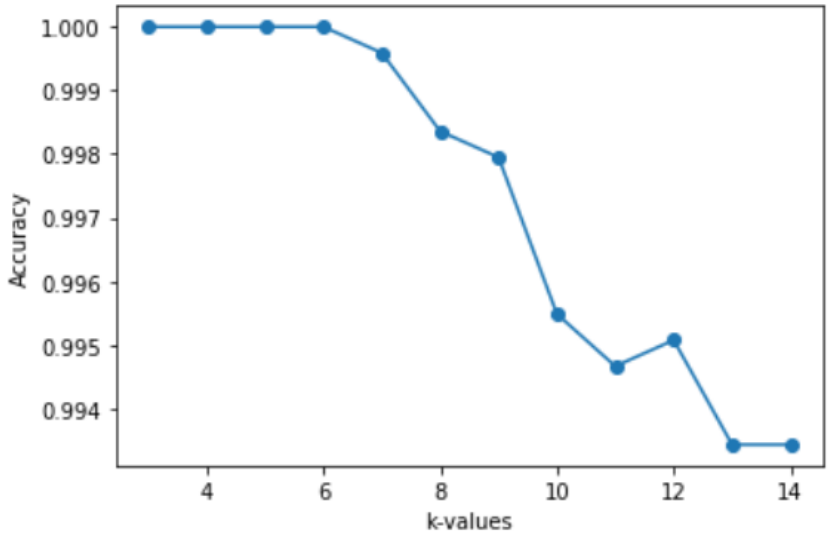




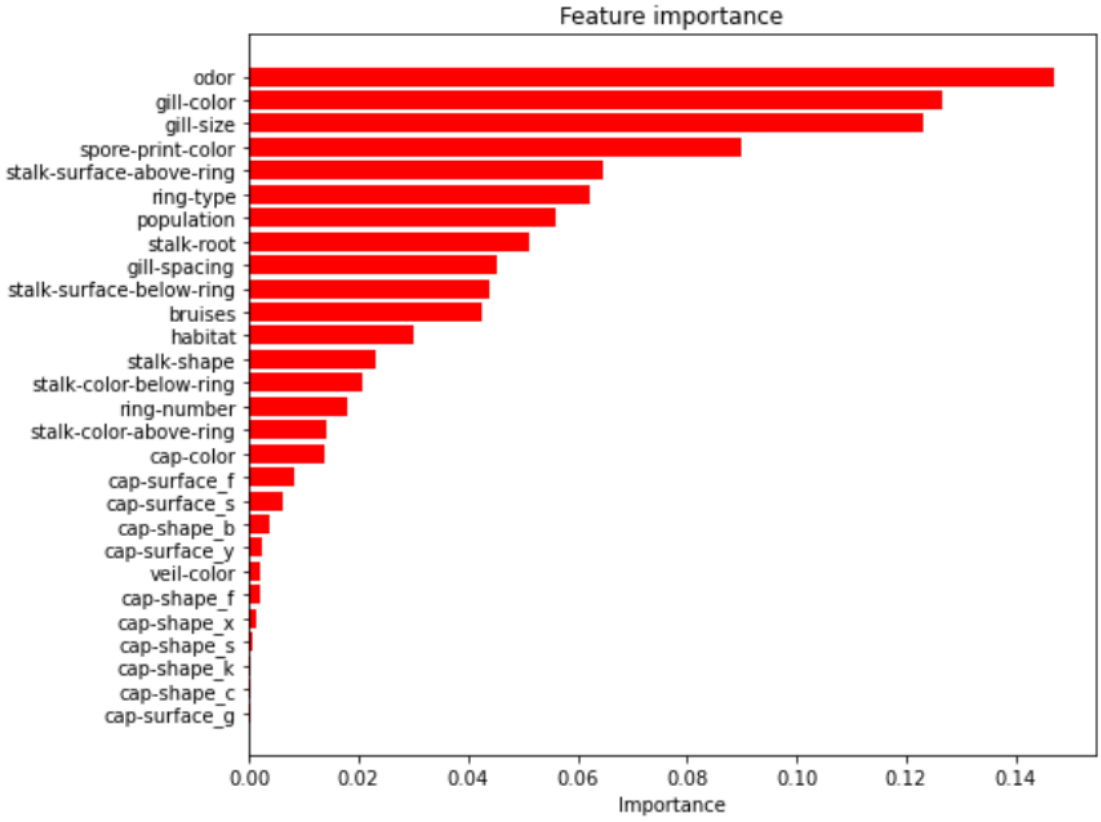
7.6 Heatmap


class
cap-color
bruises
odor
gill-spacing
gill-size
gill-color
stalk-shape
stalk-root
stalk-surface-above-ring
stalk-surface-below-ring
stalk-color-above-ring
stalk-color-below-ring
veil-color
ring-number
ring-type
spore-print-color
population
habitat
cap-shape_b
cap-shape_c
cap-shape_f
cap-shape_k
cap-shape_s
cap-shape_x
cap-surface_f
cap-surface_g
cap-surface_s
cap-surface_y
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6

7.7 Line graph



7.8 Horizontal Bar graph

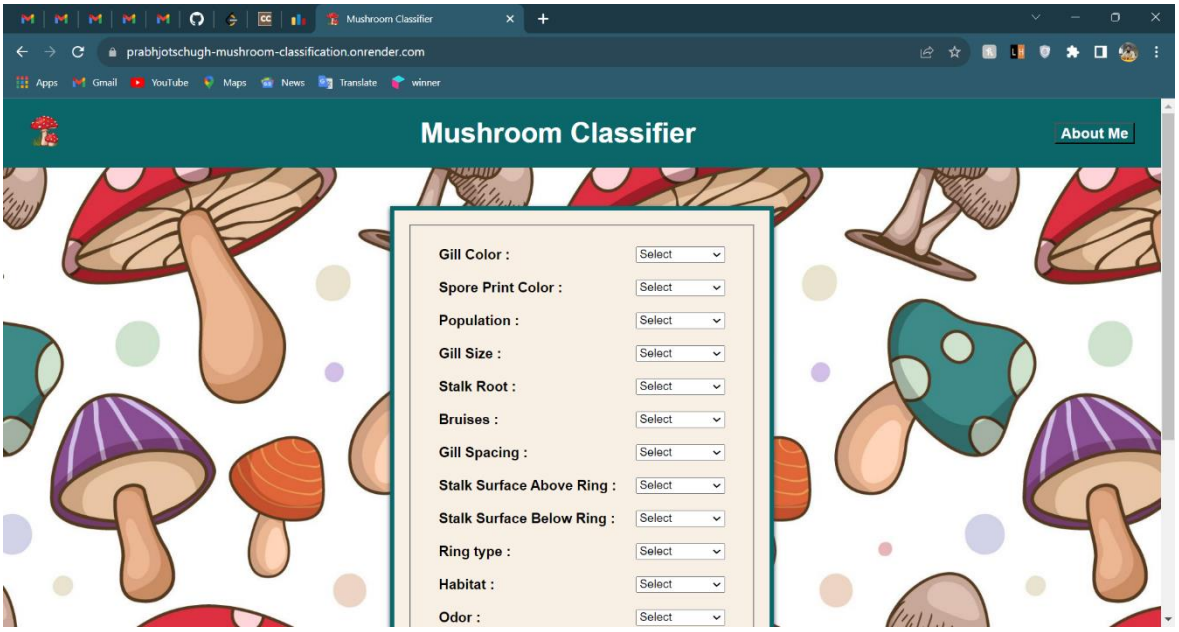


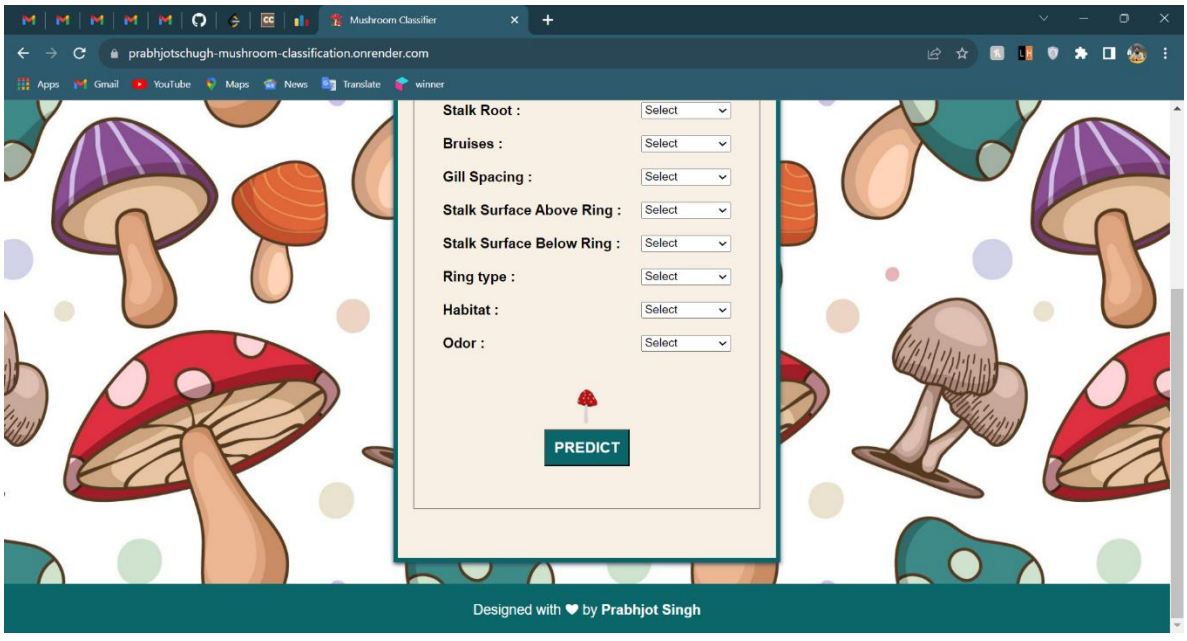
8. Web Interface

App Link - https://prabhjotschugh-mushroom-classification.onrender.com/

8.1 Home Page

When the user clicks on the app link given above, it will direct user to our home page which looks like below:





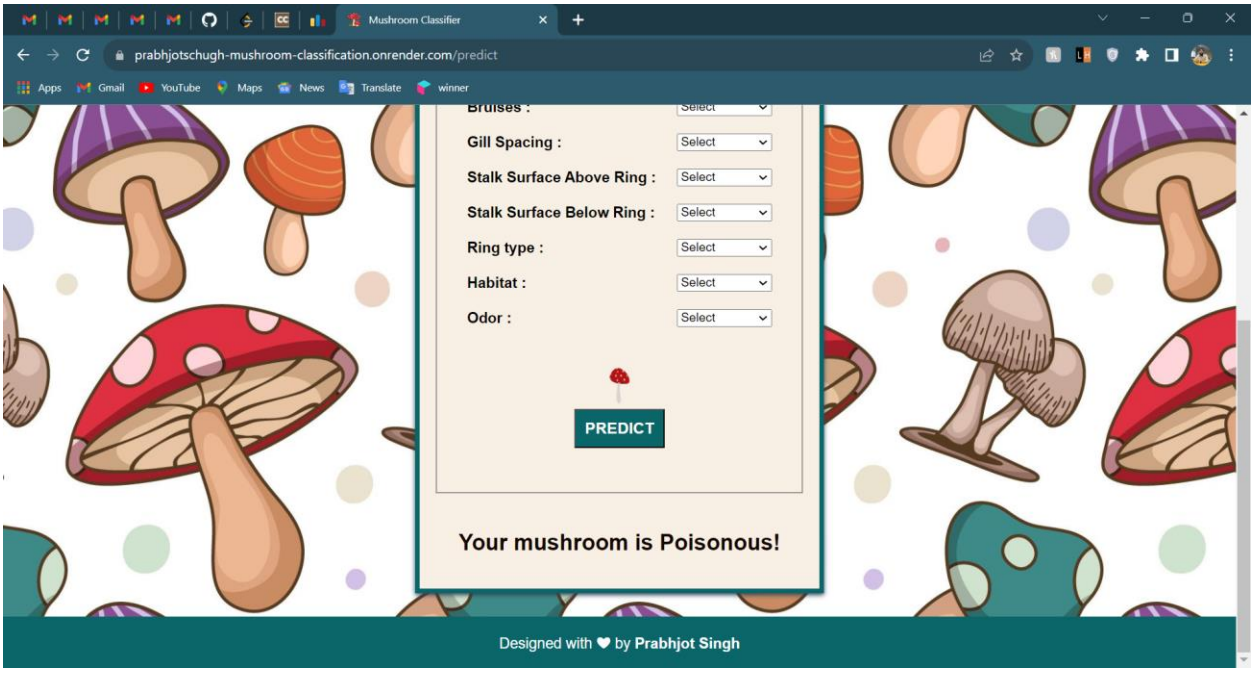
The interface presents a comprehensive set of 12 input fields, each in the form of a dropdown menu. These fields are to be carefully selected based on the specific attributes of the mushroom in question. The purpose of this selection is to determine the mushroom's classification as either edible or poisonous. Upon completing the input, a "Predict" button is provided to initiate the process. This button, when activated, leads the user to a results page that delivers a conclusive verdict on the mushroom's edibility.

The navigation bar features an "About Me" button, which serves as a direct link to my portfolio—a curated collection of my work and accomplishments. Positioned in the footer is my name, designed as a hyperlink that, when clicked, seamlessly redirects the user to my LinkedIn profile. These navigational elements collectively enhance user experience by offering quick access to additional information about me and my professional endeavors.

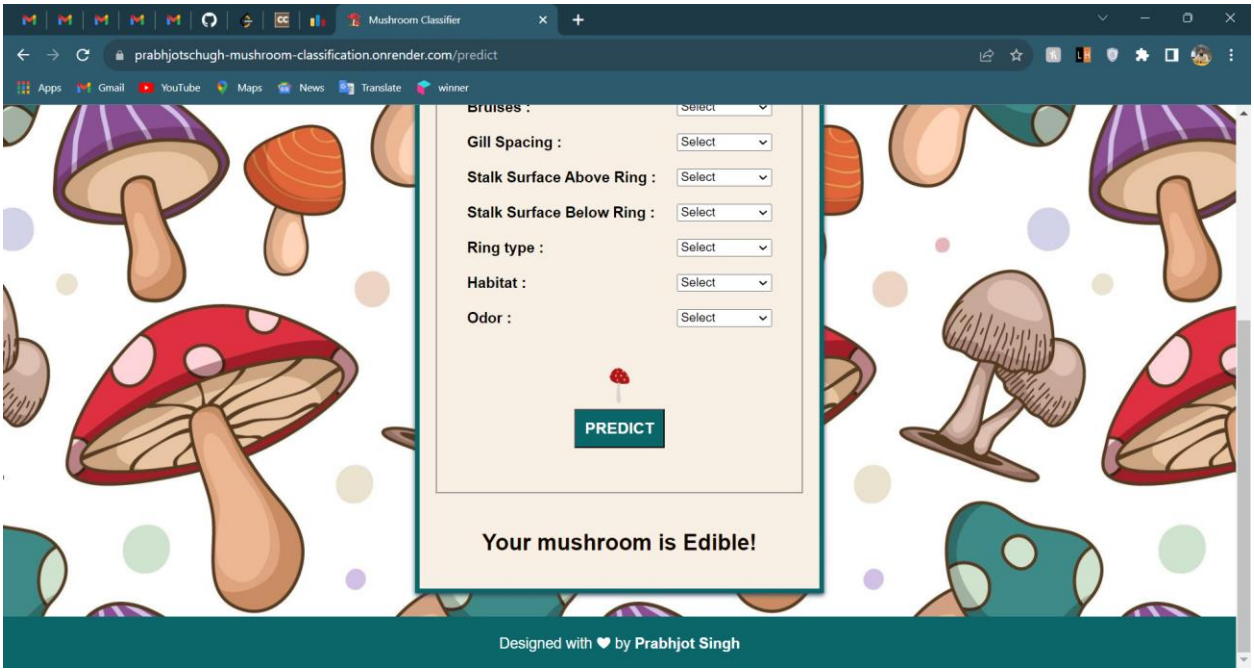
9. Sample Test Cases

In each of the ensuing examples, our focus will be directed towards the outputs that emanate from the predictive process. These outputs hold the key to determining the nature of the mushroom's classification – whether it falls under the "Edible" category, rendering it safe for consumption, or the "Poisonous" category, signifying potential harm. The robust predictive algorithm will promptly generate these outcomes based on the selected attributes, offering valuable insights into the nature of the mushroom and facilitating informed decisions about its edibility.

9.1 Poisonous Mushroom Example



9.2 Edible Mushroom Example



10. Summary

10.1 Dataset Balance:

The dataset comprises two distinct class types: 'poisonous' with 3916 instances and 'edible' with 4208 instances, reflecting a well-balanced distribution between the classes. Within the dataset:

- **Cap-Surface Variations:** Four types of 'cap-surface' exist. Notably, among the 'edible' mushrooms, the attribute 'cap-surface: g: grooves' is absent.
- **Bruises and Edibility**: The presence or absence of bruises doesn't solely determine the edibility of mushrooms.
- **Gill Spacing and Edibility:** Gill spacing, whether 'Close' or 'Crowded,' doesn't provide a definitive indicator of edibility.
- **Gill Size and Edibility:** The edibility of mushrooms is not exclusively influenced by Gill Size, whether 'Narrow' or 'Broad.'
- **Gill Color and Edibility:** 'Edible' mushrooms lack Gill Colors Buff and Green, while 'poisonous' mushrooms avoid Gill Colors Red and Orange.
- **Stalk Root and Edibility:** 'Poisonous' mushrooms do not exhibit the 'Rooted' type in the Stalk Root attribute.
- **Stalk-Surface-Above/Below-Ring and Edibility:** The presence of variations in Stalk-Surface-Above-Ring and Stalk-Surface-Below-Ring attributes does not solely determine the edibility of mushrooms.
- **Ring-Type and Edibility:** 'Edible' mushrooms exclude Ring Types 'Large' and 'None,' while 'poisonous' mushrooms avoid Ring Type 'Flaring.'
- **Spore-Print-Color and Edibility:** 'Edible' mushrooms lack Spore-Print-Color 'Green,' and 'poisonous' mushrooms avoid Spore-Print-Colors 'Purple,' 'Orange,' 'Yellow,' and 'Buff.'
- **Population Type and Edibility:** 'Poisonous' mushrooms lack Population Types 'Numerous' and 'Abundant.'
- **Habitat Type and Edibility:** 'Poisonous' mushrooms avoid Habitat Type 'Waste.'

10.2 Model Performance

The Random Forest Classifier model showcases impeccable predictive prowess, achieving a 100% accuracy on both training and test data, attesting to its reliability in distinguishing between 'poisonous' and 'edible' mushrooms.