High Level Design (HLD)

Mushroom Classification

- **Riyaz Khorasi & Sachin Shinde**

Revision Number: 2.0
Last date of revision: 20/10/2021

Document Version Control

Date Issued	Version	Description	Author
18/10/2021	1	Initial HLD — V1.0	Riyaz
19/10/2021	2	Design Details — V1.1	Sachin
20/10/2021	3	Data Visualization — V1.2	Sachin

Contents

Abstract

In recent years, the popularity of mushrooms as a superfood and the understanding of their vast health benefits has surged. What used to be typically seen as just a traditional food, mushrooms are now being widely consumed and acknowledged for their healing and health abilities. There are thousands of species of Mushrooms in the world; they are edible and non-edible being poisonous. It is difficult for non-expertise people to Identify poisonous and edible mushrooms of all species manually. So, a computer aided system with software or algorithm is required to classify poisonous and nonpoisonous mushrooms. This project is presented on classification of poisonous and nonpoisonous mushrooms. Most of the research works to classify the type of mushroom have applied, machine learning techniques like Naïve Bayes, K-Neural Network, Support vector Machine (SVM), Decision Tree techniques.

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: o
  - 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.

1.3 Definitions

Term	Description
EDA	Exploratory Data Analysis (Visualizing the data using various plots)
Bruises	Damage of plant tissue by pressure
VIF	Variance Inflation Factor (used to detect the presence of multicollinearity).
PCA	Principal Component Analysis (used for Dimensionality reduction)

2 General Description

2.1 Product Perspective

Through classification modeling, this project aims to classify mushrooms as edible or poisonous.

2.2 Problem statement

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.

2.3 PROPOSED SOLUTION

The solution proposed here is the classical machine learning tasks like Data Exploration, Data Cleaning, Feature Engineering, Model Building and Model Testing. Trying out different machine learning algorithms that's best fit for the above case to identify mushroom is poisonous or edible.

2.4 Data Requirements

Data requirement completely depend on our problem statement.

- We need data that is balanced and must have at least 5000 values.
- We require at least 18-20 variables for accurate prediction of classifying mushrooms.

2.5 Tools used

- We use Python (3.1) programming language and frameworks such as NumPy, Pandas, Scikit-learn are used to build the whole model.
- PyCharm is used as IDE.
- For visualization of the plots, Matplotlib and Seaborn are used.
- Heroku is used for deployment of the model.
- Tableau is used for dashboard creation.
- Flask is an API of Python that allows us to build up web-applications.
- Front end development is done using HTML/CSS
- GitHub is used as version control system.

2.6 Constraints

The Mushroom classification solution system must be user friendly, as automated as possible and users should not be required to know any of the workings.

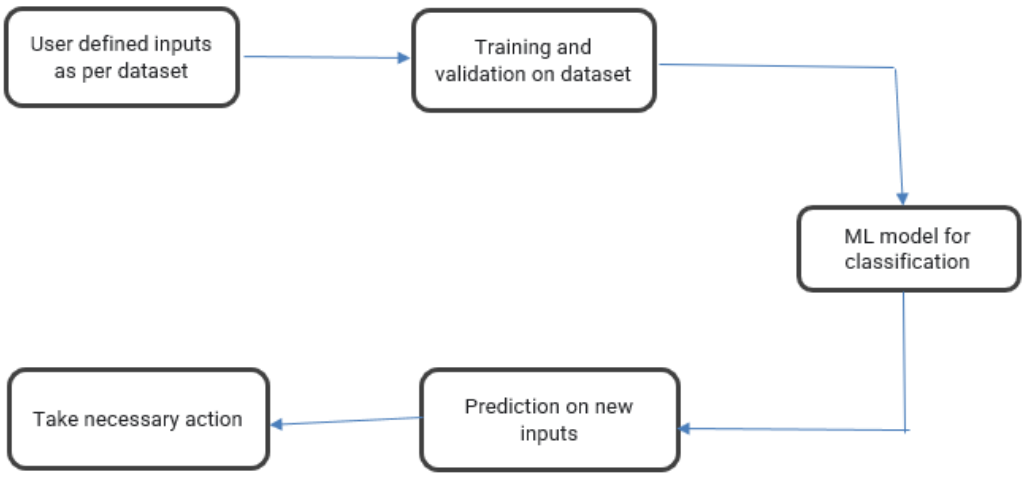
2.7 Assumptions

The main objective of the project is to implement the use cases as previously mentioned (2.2 Problem Statement) for new dataset that comes through user defined data. Machine Learning based model is used for classifying the above-mentioned use cases based on the input data. It is also assumed that all aspects of this project have the ability to work together in the way the designer is expecting.

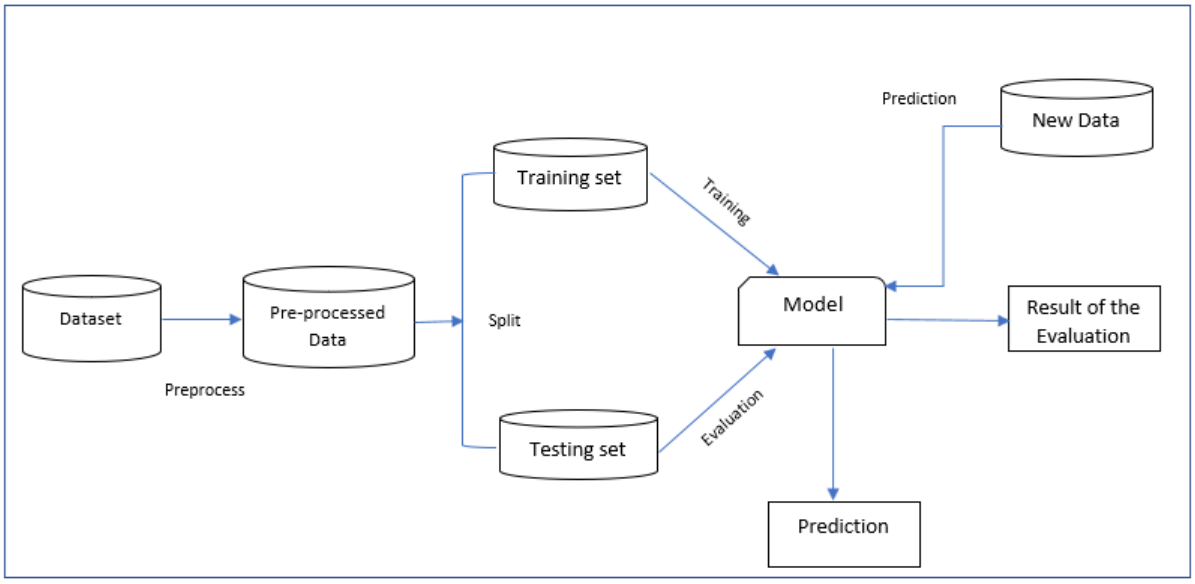
3 Design Details

3.1 Process Flow

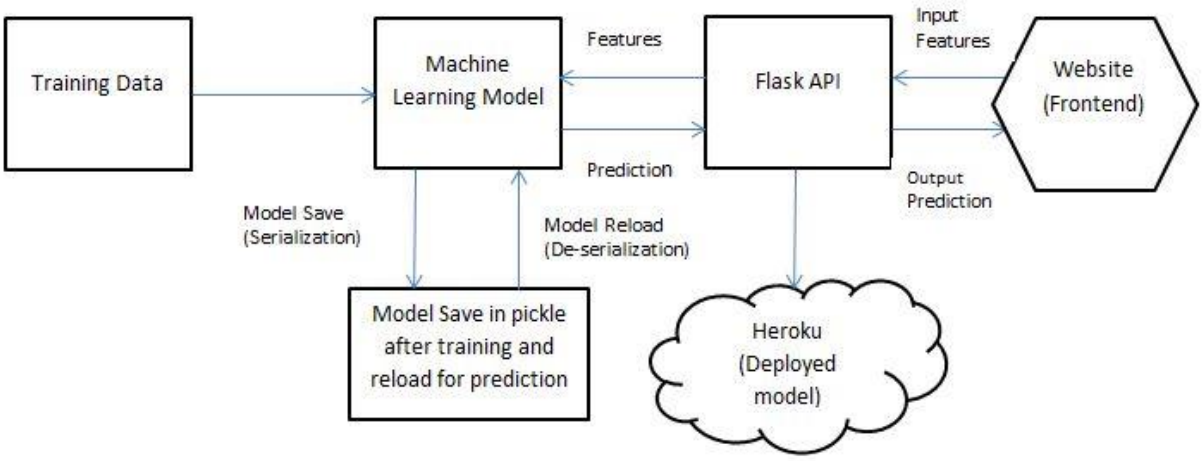
For classifying the mushroom, we will use a machine learning base model.Below is the process flow diagram is as shown below.



3.1.1 Model Training and Evaluation



3.1.2 Deployment Process



3.2 Error Handling

Should errors be encountered, an explanation will be displayed as to what went wrong? Anerror will be defined as anything that falls outside the normal and intended usage.

4 Performance

The Mushroom classification solution is used for classification of mushrooms. it will inform user whether mushrooms are edible or poisonous and takes necessary action, so it should be as accurate as possible. Also, model retraining is very important to improve the performance.

4.1 Reusability

The code written and the components used should have the ability to be reused with noproblems.

4.2 Application Compatibility

The different components for this project will be using Python as an interface between them. Each component will have its own task to perform, and it is the job of the Python to ensure proper transfer of information.

4.3 Resource Utilization

When any task is performed, it will likely use all the processing power available until that function is finished.

4.4 Deployment

Heroku: There are various strategies to deploy and run the model in the cloud platform, which seems a better option for most of the cases because of the availability of tools such as Google Cloud Platform, Azure, Amazon Web Services, and Heroku.
Heroku is a container-based cloud Platform as a Service (PaaS). Developers use Heroku to deploy, manage, and scale modern apps. This platform is elegant, flexible, and easy to use, offering developers the simplest path to getting their apps to market.
Heroku is fully managed, giving developers the freedom to focus on their core product without the distraction of maintaining servers, hardware, or infrastructure. The Heroku experience provides services, tools, workflows, and polyglot support—all designed to enhance developer productivity.

5 Data Visualizations

Data Visualizations will be implemented to display Univariate and Bivariate Analysis to obtain meaningful data insights and compare features for the analysis of classification.

6 Conclusion

Our tuned classification models all performed really well with the dataset. Random Forest, which had an accuracy score of 99% would normally be a great that the other tested models performed perfectly as well, the other models are also much better suited to classify mushrooms. Since our models performed so well, it was clear to us that they were able to identify specific traits that greatly influenced the classification of an edible versus poisonous mushroom. And that was exactly what we were hoping for!

7 References

1. https://www.kaggle.com/uciml/mushroom-classification
2. https://scikit-learn.org/stable/supervised_learning.html