



MUSHROOM CLASSIFICATION

Low Level Design Document



Prepared by
GOURAB MAHAPATRA

Abstract

The Mushroom Classifier Prediction Project aims to develop a robust machine learning model capable of accurately identifying and classifying mushroom species based on their visual features. This project is motivated by the critical need for reliable tools to differentiate between edible and poisonous mushrooms, considering the potential life-threatening consequences of consuming toxic varieties.

Table of Contents

1. Introduction	3
1.1. What is Low Level Design Document	3
1.1.1. Scope	3
2. Technical Specification	4
2.1. Dataset	4
2.1.1. Dataset Overview	4
2.1.2. Input Schema	4
2.2. Predicting Credit Fault	4
2.3. Logging	5
2.4. Deployment	5
3. Architecture	6
4. Architecture Description	7
4.1. Data Description	7
4.2. Data Exploration	10
4.3. Feature Engineering	10
4.4. Train Test Split	10
4.5. Model Building	10
4.6. Save the Model	10
4.7. Cloud Setup & Pushing the App to the Cloud	10
4.8. Application Start & Input Data by User	11
4.9. Prediction	11

1. Introduction

1.1.What is LLD?

The primary objective of this document is to provide an in-depth description of the Deep EHR System. It aims to elucidate the systems objectives and functionalities, outline its interfaces, define its core functions, address operational constraints, and detail its responses to external inputs. This document serves as a valuable resource for both stakeholders and developers involved in the systems development and is intended for submission to senior management for their approval.

1.2.Scope

The Deep EHR System is envisioned as a sophisticated web-based application strategically designed for the purpose of classifying mushrooms based on their features. This predictive tool employs a dataset containing descriptions of various mushroom species, utilizing features such as cap shape, colour, Odor, and more. The system's primary objective is to predict whether a given mushroom is edible or poisonous based on its distinctive characteristics.

2. Technical Specification

2.1.Dataset

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

Dataset Overview

The dataset 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. The Guide clearly states that there is no simple rule for determining the edibility of a mushroom; no rule like "leaflets three, let it be" for Poisonous Oak and Ivy

Input Schema

Feature Name	Datatype	Null/Required
class	Object	Required
cap-shape	Object	Required
cap-surface	Object	Required
cap-color	Object	Required
bruises	Object	Required
odor	Object	Required
gill-attachment	Object	Required
gill-spacing	Object	Required
gill-size	Object	Required
gill-color	Object	Required
stalk-shape	Object	Required
stalk-root	Object	Required
stalk-surface-above-ring	Object	Required
stalk-surface-below-ring	Object	Required

Mushroom Classification Prediction

stalk-color-above-ring	Object	Required
stalk-color-below-ring	Object	Required
veil-color	Object	Required
ring-number	Object	Required
ring-type	Object	Required
spore-print-color	Object	Required
population	Object	Required
habitat	Object	Required

2.2.Predicting Mushroom Class

- The system initiates by presenting a user interface with input fields.
- The user then provides the necessary information.
- Subsequently, the system is tasked with making a prediction regarding the likelihood of a given mushroom being poisonous or edible based on its features.

2.3.Logging

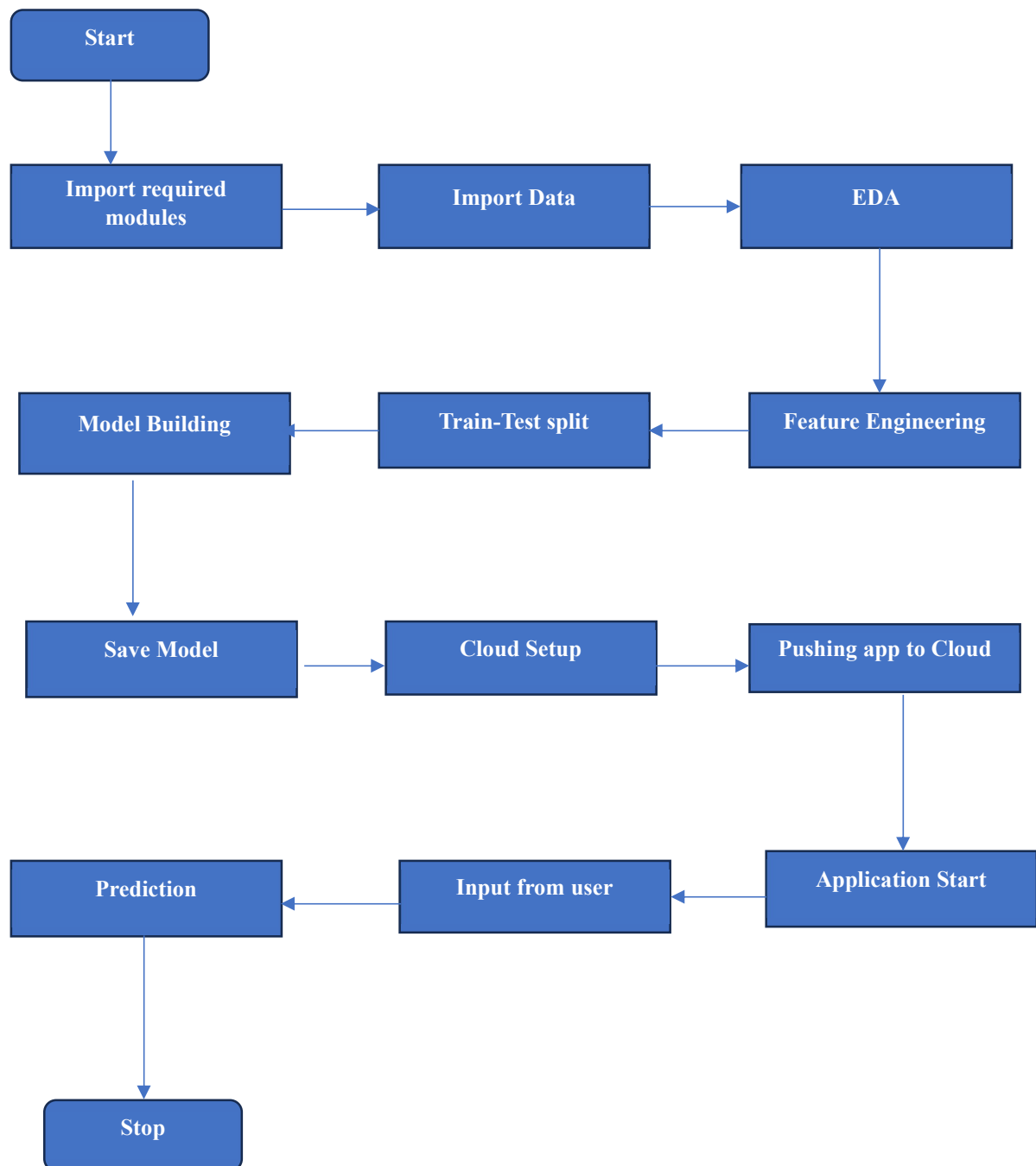
It's essential to maintain a comprehensive log of every user activity. The system autonomously recognizes the appropriate moments for logging.

- Logging is obligatory for all system processes and flows.
- Developers have the flexibility to select their preferred logging methods, whether its database logging or file logging.
- Importantly, the system must remain operational and not experience performance issues, even with extensive logging. This emphasis on logging is primarily for effective issue debugging and, thus, is a mandatory practice.

2.4.Deployment

Deployed in AWS Elastic Beanstalk

3. Architecture



4. Architecture Description

4.1.Data Description

This dataset 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.

There are 23 variables:

- class:
 - 1= edible
 - 2= poisonous
- cap_shape:
 - 1=bell
 - 2=conical
 - 3=convex
 - 4=sunken
 - 5=flat
 - 6=knobbed
- cap_surface:
 - 1=fibrous
 - 2=grooves
 - 3=scaly
 - 4=smooth
- cap_color:
 - 1=brown
 - 2=buff
 - 3=cinnamon
 - 4=gray
 - 5=green
 - 6=pink
 - 7=purple
 - 8=red
 - 9=white
 - 10=yellow
- bruises:
 - 1=no
 - 2=yes
- odor:
 - 1=almond
 - 2=anise
 - 3=creosote
 - 4=fishy
 - 5=foul
 - 6=musty
 - 7=none
 - 8=pungent

Mushroom Classification Prediction

- 9=spicy
- gill_attachment:
 - 1 = attached
 - 2 = free
- gill_spacing:
 - 1=close
 - 2=crowded
- gill-size:
 - 1=broad
 - 2=narrow
- gill_color:
 - 1=black
 - 2=brown
 - 3=buff
 - 4=chocolate
 - 5=gray
 - 6=green
 - 7=orange
 - 8=pink
 - 9=purple
 - 10=red
 - 11=white
 - 12=yellow
- stalk_shape:
 - 1=enlarging
 - 2=tapering
- stalk_root:
 - 1=bulbous
 - 2=club
 - 3=equal
 - 4=rooted
- stalk-surface-above-ring:
 - 1=fibrous
 - 2=scaly
 - 3=silky
 - 4=smooth
- stalk-surface-below-ring:
 - 1=fibrous
 - 2=scaly
 - 3=silky
 - 4=smooth
- stalk-color-above-ring:
 - 1=brown
 - 2=buff
 - 3=cinnamon
 - 4=gray

Mushroom Classification Prediction

- 5=orange
 - 6=pink
 - 7=red
 - 8=white
 - 9=yellow
- stalk-color-below-ring:
 - 1=brown
 - 2=buff
 - 3=cinnamon
 - 4=gray
 - 5=orange
 - 6=pink
 - 7=red
 - 8=white
 - 9=yellow
- Veil-color:
 - 1=brown
 - 2=orange
 - 3=white
 - 4=yellow
- ring-number:
 - 1=none
 - 2=one
 - 3=two
- ring-type:
 - 1=evanescent
 - 2=flaring
 - 3=large
 - 4=none
 - 5=pendant
- spore-print-color:
 - 1=black
 - 2=brown
 - 3=buff
 - 4=chocolate
 - 5=green
 - 6=orange
 - 7=purple
 - 8=white
 - 9=yellow
- population:
 - 1=abundant
 - 2=clustered
 - 3=numerous
 - 4=scattered
 - 5=several

Mushroom Classification Prediction

- 6=solitary
- habitat:
 - 1=grasses
 - 2=leaves
 - 3=meadows
 - 4=paths
 - 5=urban
 - 6=waste
 - 7=woods

4.2.Data Exploration

We conduct a detailed exploration for each feature (categorical type), one at a time. Within each type, we systematically examine, visualize, and analyze each variable individually, documenting our findings. Additionally, we may make minor modifications to the data, such as renaming columns for improved clarity and ease of understanding.

4.3.Feature Engineering

Categorical variables have been encoded to facilitate data analysis and modelling

4.4.Train Test Split

The dataset has been divided into two subsets: a training set, which comprises 70% of the data, and a test set, which consists of the remaining 30%. This split allows for training and testing machine learning models.

4.5.Model Building

Several models have been constructed, and the dataset has been used to train and evaluate these models. The performance of each model has been thoroughly compared, and the best-performing model has been selected based on various evaluation metrics and criteria.

4.6.Save The Model

The selected model has been saved by converting it into a pickle file. This allows for easy storage and retrieval of the model for future use.

4.7.Cloud Setup & Pushing the App to The Cloud

AWS (Amazon Web Services) has been chosen as the deployment platform for the application. The application files have been loaded from the GitHub repository to the AWS environment, ensuring that the application is hosted and accessible on AWS infrastructure.

Mushroom Classification Prediction

4.8.Application start & input data by user

The application has been initiated and is now ready for use. You can enter the required inputs into the application to perform the desired tasks.

4.9.Prediction

Once the user submits the features of a mushroom as an input, the mushroom classification application will execute the trained model to generate predictions. The output will be presented as a message, conveying information about whether the submitted mushroom is likely to be poisonous or edible based on its features.