Architecture

Mushroom Classification

Written by	Chetan Ullagaddi
Document version	1.0
Last Revision Date	26/09/2021

Document Control

Date	Version	Author	Comments
26/09/2021	Version-1.0	Chetan Ullagaddi	Introduction and Architecture Defined

Approval Status

Version	Review Date	Reviewed By	Approved By	Comments
Version-1.0				

Contents

Page no.

1. Introduction

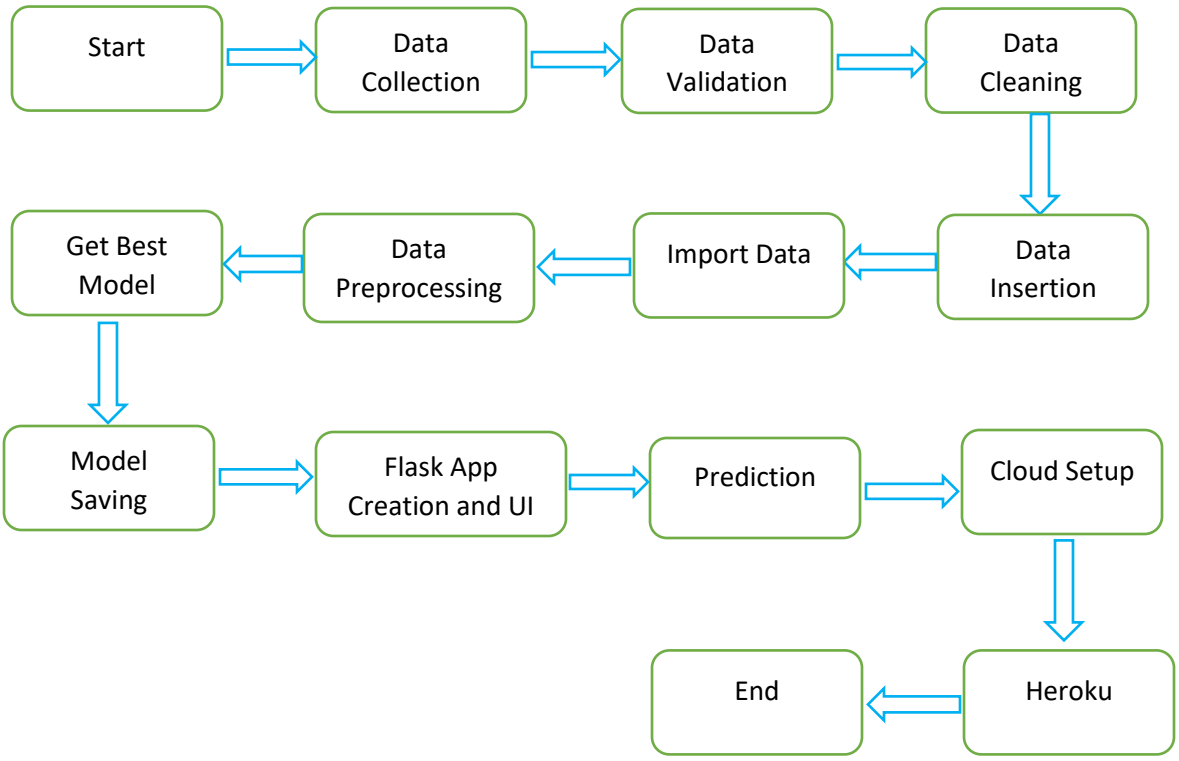
1.1 Why this Architecture Design Document?

The purpose of this document is to provide a detailed architecture design of the Mushroom Classification by focusing of four key quality attributes:

Usability, availability, maintainability, testability.

This document will address the background for this project, and the architecturally significant function requirements. The intention of this document is to help the development team to determine how the system will be structured at the highest level. Finally, the project coach can use this document to validate that the development team is meeting the agreed-upon requirements during the evaluation of the team's efforts.

1 Architecture



2 Architecture Description

3.1 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.

3.2 Import Data

Data Import from Database - The data in a stored database is imported as a CSV file to be used for Data Pre-processing and Model Training.

3.3 Data Cleaning

There are no null values in the data and all the variables are categorical, some of the observations meaningless and they are converted into meaningful observation.
Example: "?" is converted into letter "m" (Missing)

3.4 Exploratory Data Analysis

EDA is done in such way that every independent variable in the dataset is visualized by multiple bar plot with respect to dependent variable having classes of poisonous and edible mushrooms.

3.5 Data Preprocessing

Data preprocessing steps are converting categorical variables into numerical variables using label encoding method and train and test split of the data etc.

3.6 Model Building

After Data preprocessing split the data train and test (Simple Random Sampling) and implemented different Classification Machine Learning Algorithm. Random Forest model gives the better accuracy.

3.7 Model Dump

After comparing all accuracies and finding the best model for the dataset I have created a model and dumped the model in a pickle file format with the help of pickle module.

3.8 Data from User

Here The user will have to enter all the features name in correct order and have to submit it to the model with the help of UI interface. The data will be fed to the model which will predict whether the feature set represents a mushroom is edible or not.

3.9 Data Validation

Here Data Validation will be done, given by the user.

3.10 Model Call for specific input

Based on the User Input will be thrown to the backend in the variable format then it will be converted into a NumPy array which will be fed to our model. The loading of the pickle file will be done and then the model will predict whether the Input is edible or not by sending the result to our html page.

3.11 User Interface

In frontend creation I have made a user interactive page where users can enter their input values to our application. In their frontend page I have made a form which has beautiful style with CSS. This HTML user input data is transferred in variable format to the backend. Made these html fully in a decoupled format.

3.12 Deployment

I will be deploying the model with the help of Heroku cloud platform.