

**Mushroom Type Prediction**

**Project Architecture**

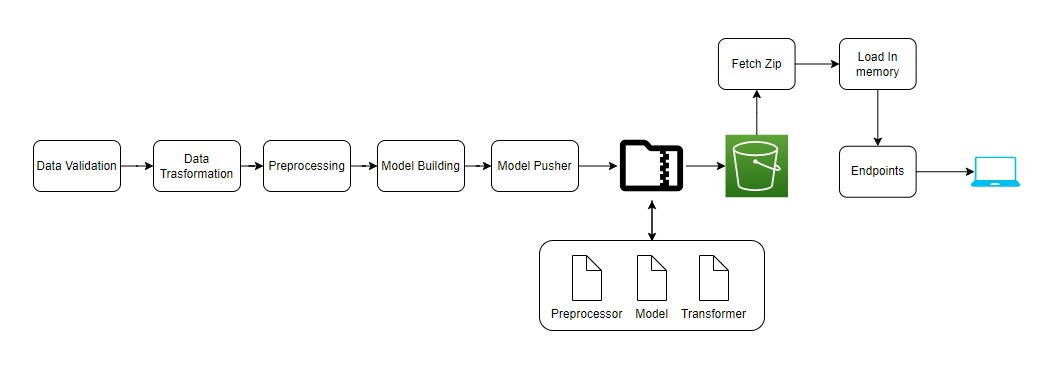
# Domain : Machine Learning

# Creator : Lalatendu Dalai

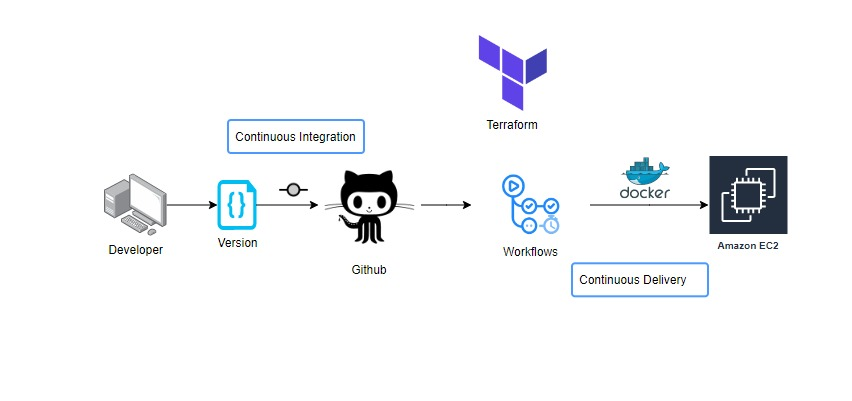
# Date : 11.04.2023

# Architecture

# Project Architecture



# Deployment Architecture



# Architecture Description

# Data Preparation

# Data Description

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.

# Data Preprocessing

In data preprocessing step, we check if there missing data, duplicate values, and datatypes of each feature. In our dataset, there was not any null and duplicate values .

# Exploratory Data Analysis

This step includes bivariate and univariate analysis of features. Checking outliers using boxplots, and outlier treatment is carried out as well. Distribution of the features are plotted to see to what extent our data is skewed.

# Feature Engineering

In this part, the datatypes of the features were checked whether it belongs same datatypes or different datatypes. Outliers were checked using boxplot but there is no such major outliers in the dataset.

# Model Development

# Model Implementation

After train and test splitting, pipeline containing Standard Scaler and Label Encoder was fitted to several models such as RandomForest Classifier,XGB Classifier, KNeighbors Classifier, etc. Their f1 score were obtained and it was determined that KNeighbors Classifier performs better than other models.

# Model Evaluation

Test dataset is used to evaluate the model. 20% of dataset was separated for testing. Predicted results of the model are compared with the actual data to check the amount of error. As there was no considerable change after , it helped us to overcome overfitting and perform better on new data.

# Deployment

# Designing UI

For this project, a user interface is built using Apache Air Flow.It is a open-source platform for developing, scheduling, and monitoring batch-oriented workflows.

# Designing a server

A server should be created to run the UI application continuously. Amazon Web Service EC2 instance is used to create a virtual server for the application.Elastic Compute Cloud (EC2) is a virtual server in AWS for running applications on the AWS infrastructure.

# Code deployment on cloud

The codes for this machine learning model should be deployed to the cloud, so that when data is entered into the application, our code runs, and a user gets the result online.

# Deployment Process

In this stage, we containerized the code using Docker and will be deploying the model to AWS.This is a workflow diagram for the Recipe Recommendation..