MUSHROOM CLASSIFICATION

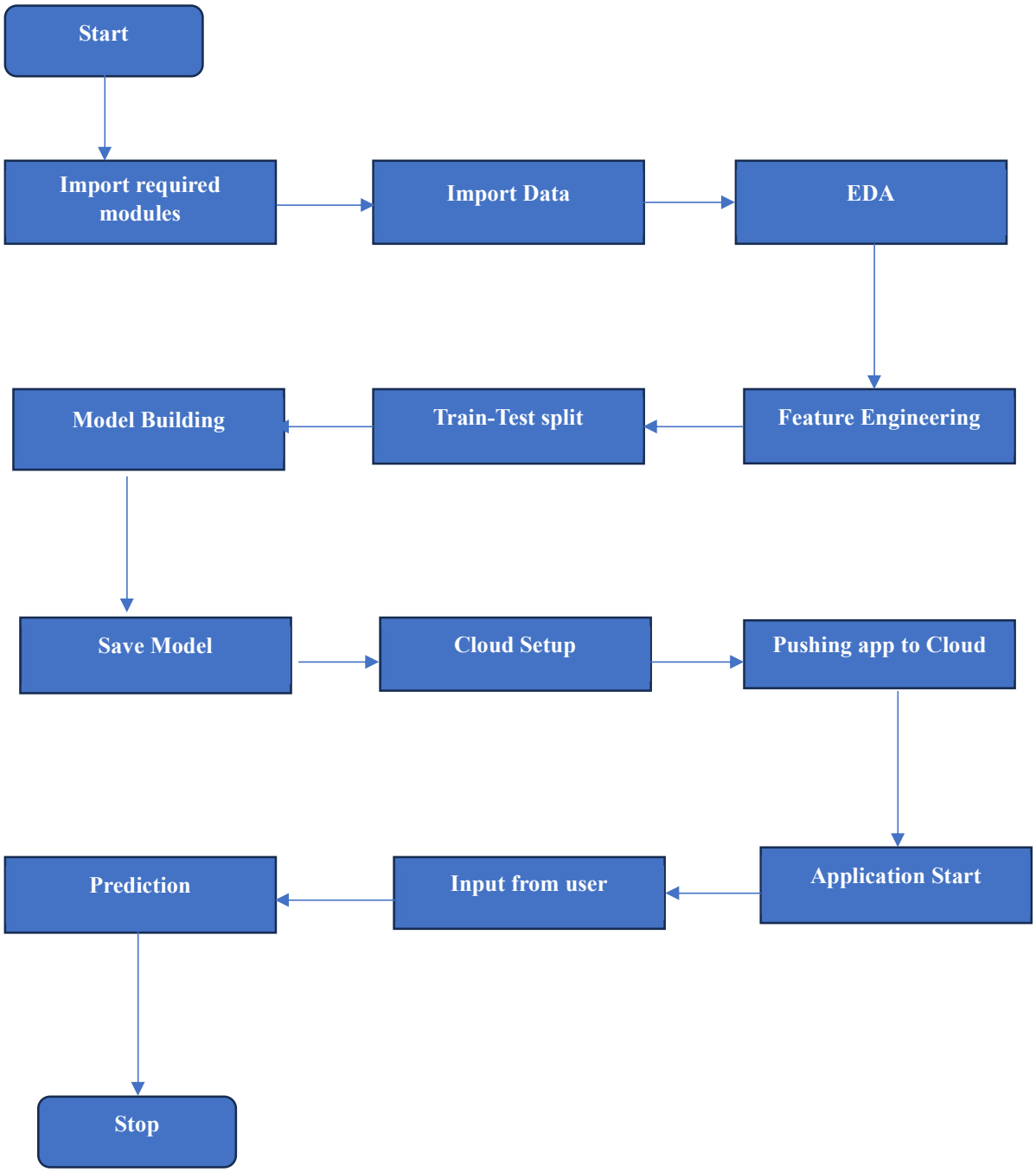
Architecture Document

PREPARED BY
GOURAB MAHAPATRA

Table of Contents

Architecture

1. Architecture



2. Architecture Description

2.1. 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
    - 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

Architecture

    - 4=gray
    - 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

Architecture

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

2.2. 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.

2.3. Feature Engineering

Categorical variables have been encoded to facilitate data analysis and modelling

2.4. 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.

2.5. 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.

2.6. 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.

2.7. 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.

2.8. 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.

2.9. 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.