High-Level Design (HLD) (Mushroom Classification)

Contents

Abstract

Mushrooms are an essential nutrient-dense element in our diet. Because most mushrooms are toxic (inedible), we must distinguish between poisonous and edible mushrooms. Machine learning (ML) techniques such as logistic regression, decision trees, Random forest, and others were used to categorize mushroom attributes as edible or not. There has been little study on mushroom classification; existing research has concentrated on applying ML approaches separately, with some algorithms outperforming others in terms of accuracy. This study developed an integrated model that combines the most accurate technique's judgements into a single decision rather than addressing them separately. Kaggle mushroom dataset downloaded. The results demonstrate that the xgboost regressor outperforms other strategies by 100% accuracy score.

1 Introduction

1.1 Why this High-Level Design Document?

The goal of this High-Level Design (HLD) Document is to supplement the current project description with the required depth to represent an acceptable model for coding. This paper is also designed to aid in the detection of conflicts prior to coding and to serve as a reference handbook 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:
    - Security
    - Reliability
    - Maintainability
    - Portability
    - Reusability
    - Application compatibility
    - Resource utilization
    - Serviceability

1.2 Scope

The HLD documentation describes the system's structure, including the database architecture, application architecture (layers), application flow (Navigation), and technology architecture. The HLD employs non-technical to moderately technical language that should be intelligible to system administrators.

2. General Description

2.1 Product Perspective

The Mushroom prediction solution system is a machine learning model based on data science that assists us in detecting edible and toxic mushrooms and taking appropriate action.

2.2 Problem Statement

The Audubon Society Field Guide to North American Mushrooms includes descriptions of hypothetical samples belonging to 23 Agaricus and Lepiota Family Mushroom species (1981). Each species is classified as certainly edible, definitely toxic, or maybe edible but not advised. This last category has been combined with the poisonous category. The Guide claims explicitly that there is no straightforward criterion for determining the edibility of a mushroom, such as "leaflets three, leave it alone" for Poisonous Oak and Ivy. The main objective is to figure out which mushrooms are harmful and which are edible.

2.3 Proposed Solution

To address the issue, we developed a User Interface for selecting user input to forecast if the mushroom is toxic or edible using our trained ML model. After processing the input, the model's final output (predicted value) is conveyed to the User.

2.4 About Dataset

Attribute Information: (classes: edible=e, poisonous=p)

- cap-shape: bell=b,conical=c,convex=x,flat=f, knobbed=k,sunken=s
- cap-surface: fibrous=f,grooves=g,scaly=y,smooth=s
- cap-color:brown=n,buff=b,cinnamon=c,gray=g,green=r,pink=p,purple=u, red=e,white=w,yellow=y
- bruises: bruises=t,no=f
- odor:almond=a,anise=l,creosote=c,fishy=y,foul=f,musty=m,none=n,punge nt=p,spicy=s
- gill-attachment: attached=a,descending=d,free=f,notched=n
- gill-spacing: close=c,crowded=w,distant=d
- gill-size: broad=b,narrow=n
- gill-color:black=k,brown=n,buff=b,chocolate=h,gray=g, green=r, orange=o, pink=p, purple=u, red=e, white=w, yellow=y
- stalk-shape: enlarging=e,tapering=t
- stalk-root: bulbous=b,club=c,cup=u,equal=e,rhizomorphs=z,rooted=r,missing=?
- stalk-surface-above-ring: fibrous=f,scaly=y,silky=k,smooth=s
- stalk-surface-below-ring: fibrous=f,scaly=y,silky=k,smooth=s
- stalk-color-above-ring: brown=n,buff=b,cinnamon=c,gray=g,orange=o,pink=p,red=e,white=w,yell ow=y
- stalk-color-below-ring: brown=n,buff=b,cinnamon=c,gray=g,orange=o,pink=p,red=e,white=w,yell ow=y
- veil-type: partial=p,universal=u veil-color: brown=n,orange=o,white=w,yellow=y
- ring-number: none=n,one=o,two=t
- ring-type: cobwebby=c,evanescent=e,flaring=f,large=l,none=n,pendant=p,sheathing= s,zone=z
- spore-print-color: black=k, brown=n, buff=b, chocolate=h, green=r, orange=o, purple=u, white= w, yellow=y
- population: abundant=a,clustered=c,numerous=n,scattered=s,several=v,solitary=y

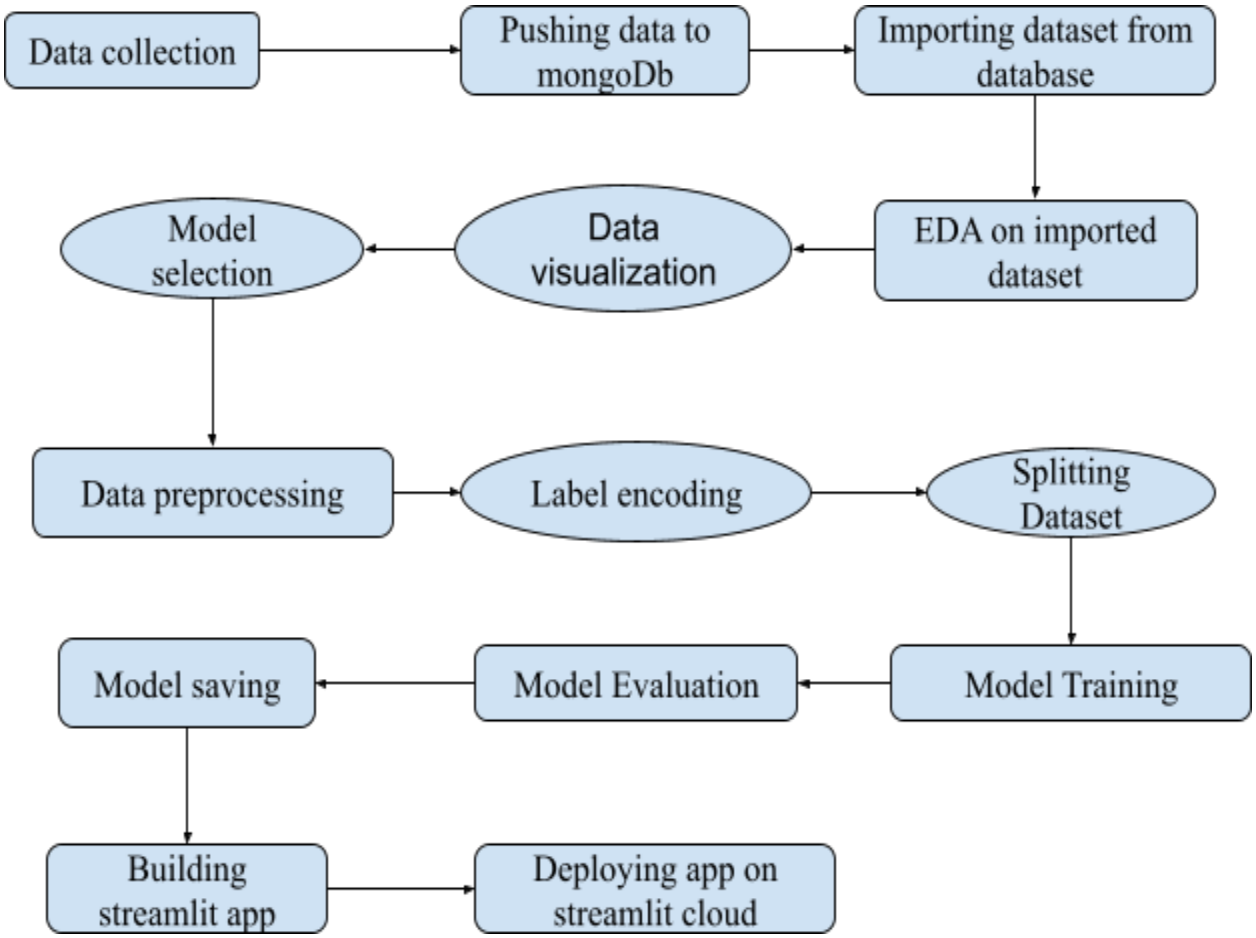
- habitat: grasses=g,leaves=l,meadows=m,paths=p,urban=u,waste=w,woods=d

2.5 Tools Used (Software Requirements)

- Python
- Streamlit
- MongoDB
- Vs Code
- Numpy
- Pandas
- Matplotlib
- Jupyter
- Sklearn

3. Design Details

3.1 Process Flow



3.2 Event log

When an error or an exception occurs, the event is logged into the system log file along with the reason and timestamp. This assists the developer in debugging system issues and correcting errors.

3.3 Error Handling

If a mistake occurs, an explanation of what went wrong will be presented. An error is defined as anything that deviates from the regular and intended usage.
Whenever the code throws an error, the traceback will show what went wrong in the code. Python traceback includes useful information that can assist you in determining what is wrong with the code. These tracebacks may appear tedious at first, but after you break them down to discover what they're trying to teach you, they may be quite useful.

4. Performance

The machine learning-based Mushroom Prediction system will be used to detect poisonous mushrooms. As a result, ASP will take the required measures. Model retraining is also essential for improving performance.

5. Deployment

A web app built with streamlit was used to deploy the model to the streamlit cloud.

6. Conclusion

This project is about pre-processing procedures, processes to discover essential features that aid in the categorization of edible and deadly mushrooms, and a comparison of attribute selection methods to determine whether both approaches yield the same outcome.