Low Level Design (HLD)

Mushroom Classifications

Written By Rohan Kadam & Vijay Kamble
Document version 1.4
Revised Date 19/09/2023

Document Control :

Date	Version	Description	Author
12/09/2023	1.0	Abstract, Introductions, Scope	Rohan Kadam
14/09/2023	1.2	Method	Vijay Kamble
17/09/2023	1.3	Tool , Deployment	Rohan Kadam
19/09/2023	1.4	Report	Vijay Kamble

Contents

Introduction

What is a Low Level Design Document?

The goal of the Low-level design document (LLDD) is to give the internal logic design of the actual program code for the Heart Disease Diagnostic Analysis dashboard. LLDD describes the class diagrams with the methods and relations between classes and programs specs. It describes the modules so that the programmer can directly code the program from the document.

What is Scope?

Low-level design (LLD) is a component-level design process that follows a step-by-step refinement process. The process can be used for designing data structures, required software architecture, source code and ultimately, performance algorithms. Overall, the data organization may be defined during requirement analysis and then refined during data design work.

Project Introduction

Data mining is also usually mentioned as knowledge discovery from Data (KDD). The aim of data mining is to mine useful and relevant information from huge databases or data warehouses. Knowledge discovery is a collaborative process, comprising of developing an understanding of the application domain, choosing and making a data set, pre-processing, and data transformation. Mushrooms live in several habitats like above the ground, on the ground, or even on the plants such as deceased wood. It contains high protein, vitamins, minerals, and antioxidants. In the science field, it is one type of fungus. Mushrooms are the most sustainably produced food not only taste but also have an excessive nutritional value. Mushrooms are also said to be a mediator

Problem Statement

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. TheGuide asserts unequivocally that there is no simple rule for judgingamushroom's edibility, such as "leaflets three, leave it be" for PoisonousOakandIvy. The main goal is to predict which mushroom is poisonous & which is edible.

Dataset Information

The dataset used in this project contains 8124 instances of mushrooms with 23 features like cap-shape, cap-surface, cap-color, bruises, odor, etc.

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

odour: almond=a,anise=l,creosote=c,fishy=y,foul=f,musty=m, none=n,pungent=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,yellow=y

stalk-color-below-ring:
brown=n,buff=b,cinnamon=c,gray=g,orange=o,pink=p,
red=e,white=w,yellow=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,pe
ndant=p,sheathing=s,zone=z

spore-print-color :
black=k,brown=n,buff=b,chocolate=h,green=r,orange=
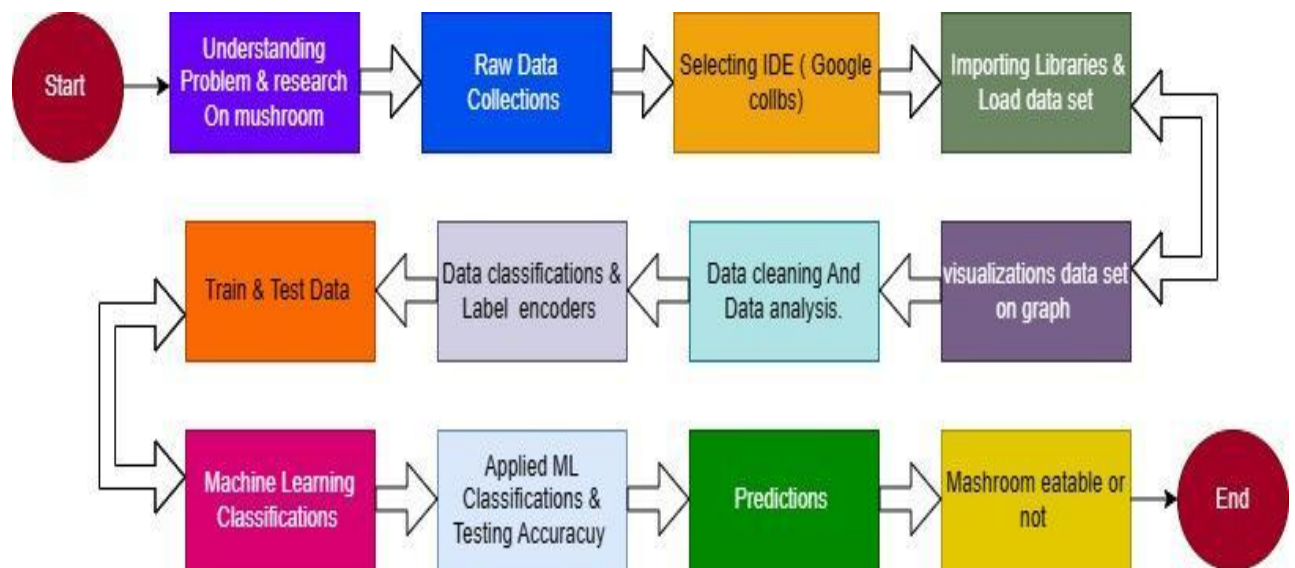o,purple=u,white=w,yellow=y

population:
abundant=a,clustered=c,numerous=n,scattered=s,severa
l=v,solitary=y

habitat:

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

Architecture

Architecture Description

1. Raw Data Collection -The Dataset was taken from iNeuron's Provided Project Description Document.
https://astra.dev/ineuron
Data-Set link :
https://www.kaggle.com/datasets/uciml/mushroom-classification

2. Data Collection and Exploration:

Obtain a dataset of mushroom samples, including features like cap shape, cap color, gill colour, odour, etc. Explore the dataset to understand its structure, check for missing values, and ensure data quality.

3. Data Preprocessing:

Handle missing data, either by imputation or removing incomplete samples.
Encode categorical features using techniques like one-hot encoding or label encoding.
Split the dataset into training and testing sets for model evaluation.

4. Model Selection:

Choose appropriate machine learning algorithms for classification, such as Decision Trees, Random Forests, Support Vector Machines, or Gradient Boosting. Consider using ensemble methods for improved performance.

5. Model Training:

Train the selected models on the training dataset using cross-validation techniques like k-fold cross-validation to tune hyperparameters.
Evaluate the models' performance on the validation set.

6. Model Evaluation:

Use various evaluation metrics such as accuracy, precision, recall, F1-score, and confusion matrix to assess model performance.
Address any issues like class imbalance or overfitting.

7. Reporting

Reporting is a most important and underrated skill of a data analytics field. Because being a Data Analyst you should be good in easy and self explanatory report because your model will be used by many stakeholders who are not from technical background.

a) High Level Design Document (HLD)

b) Low Level Design Document (LLD)

c) Architecture

d) Wireframe

e) Detailed Project Report

f) Powerpoint Presentation