Architecture Design

Mushroom Classification

Revision Number: 2.0

Last Date of revision: 08/12/2022

Document Version control

Date issued	Version	Descriptions	Author
20/11/2022	1.0	Introduction, What is an Architecture Design Document?,What is Scope?	Varun Salunkhe
30/11/2022	1.1	Technical specifications, About Dataset, Content, About this File	Saurabh Jumnalkar
05/12/2022	1.2	Logging, Technology stack	Saurabh Jumnalkar
08/12/2022	1.3	User I/O workflow	Sourabh Hawale

Contents

ABSTRACT

Mushroom is found to be one of the best nutritional foods with high proteins, vitamins, and minerals. It contains antioxidants that prevent people from heart disease and cancer. Around 45000 species of mushroom are found to be existing in worldwide. Among these, only some of the mushroom varieties were found to be edible. Some of them are really dangerous to consume. In order to distinguish between the edible and poisonous mushrooms in the mushroom dataset which was obtained from UCI Machine Learning Repository, some data mining techniques are used. Weka is a data mining tool with various machine learning algorithms that can pre-process, analyze, classify, visualize and predict the given data. Thus, to select the attributes that help better classify mushrooms, the Wrapper method and Filter method in Weka is used to identify the best attributes for the classification. The attributes 'odor' and 'spore_print_color' were chosen to be the best ones that contributed to the better classification of edible and poisonous mushrooms. After identifying the key attributes, classification is performed, a decision tree is constructed based on those attributes, and its Precision, Recall, and F-Measure values are analyzed.

1. INTRODUCTION

1. What is an Architecture Design Document?

Any software needs an architectural design to represent the design of the software, IEEE defines architectural design as "the process of defining a collection of hardware and software components and their interfaces to establish the framework for the development of a computer system." The software that is built for computer-based systems can exhibit one of these many architectures.

Each style will describe a system category that consists of:

- A set of components (ex: a database, computational modules) that will perform a function required by the system.
- The set of connectors will help in coordination, communication, and cooperation between the components.
- Conditions that how components can be integrated to form the system.
- Semantic models help the designer to understand the overall properties of the system.

2. What is Scope?

Architecture Design Document (ADD) is an architectural 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 design principles may be defined during requirement analysis and then refined during architectural design work.

2. TECHNICAL SPECIFICATIONS

2.1 About Dataset

Although this dataset was originally contributed to the UCI Machine Learning repository nearly 30 years ago, mushroom hunting (otherwise known as "shrooming") is enjoying new peaks in popularity. Learn which features spell certain death and which are most palatable in this dataset of mushroom characteristics. And how certain can your model be?

2.2 Content

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. This latter class was combined with the poisonous one. The Guide clearly states that there is no simple rule for determining the edibility of a mushroom; no rule like "leaflets three, let it be" for Poisonous Oak and Ivy.

2.3About this File

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,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,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

2.4 Logging

We should be able to log every activity done by the user.

- The System identifies at what step logging required
- The System should be able to log each and every system flow.
- Developers can choose logging methods. You can choose database logging/File logging as well.
- The system should not be hung even after using so many loggings. Logging is just because we can easily debug issues so logging is mandatory to do.

3. TECHNOLOGY STACK

Front End	Html, CSS, Javascript
Backend	Flask
Database	MongoDB
Deployment	Local host

4. USER I/O WORKFLOW