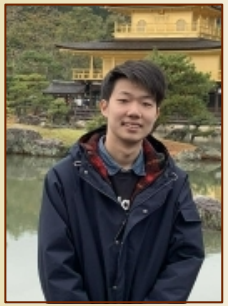
Mushroom Toxicity Identification

Sino Dragon

Table of contents

01 Our Team

Our Team



Changhao Zhang

Ningxin Kong

Xusen Shi

Yichao Ma

Yifan Han

Hongxuan Zhao

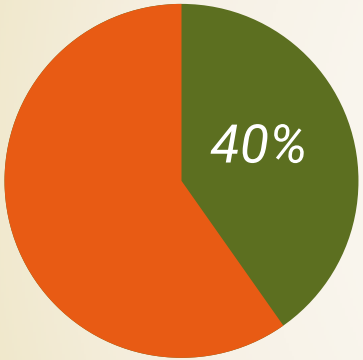
Yuheng Bai

Business

02

Poisoning Incidents

Number of accidents: 10,036

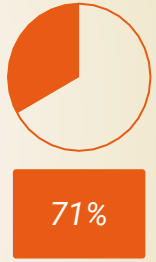


Yunnan

Yunan had 4,010 accidents, and 40% of whole country

Number of persons: 38676

Lethality rate



Incidence Rate

Our Client

· Collaboration with the Yunnan government
· Improvement of villagers' ability to recognize the toxicity of mushrooms and reduction of accidents

Strategy

01

Convenience

Application Development

02

Versatility

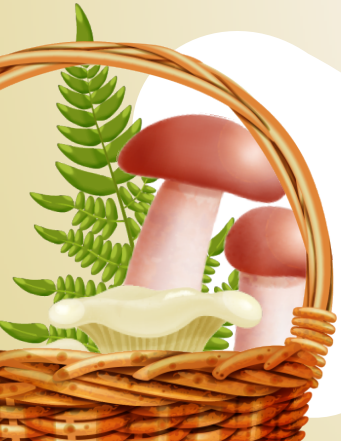
Based on signs observed by people

03

Results

Provides results for specific mushroom

03 Data Explanatory



Mushroom Data Overview

Source: "The Audubon Society Field Guide to North American Mushrooms" (1981)

Features: Mushroom features such as cap shape, cap surface, gill shape, stalk shape, and many others. All features are *categorical*

Classes:

- Edible
- Poisonous

	class	cap-shape	cap-surface	cap-color	bruises	odor	gill-attachment	gill-spacing	gill-size	gill-color	...	stalk-surface-below-ring	stalk-color-above-ring	stalk-color-below-ring	veil-type	veil-color	ring-number	ring-type	spore-print-color	popula
0	p	x	s	n	t	p	f	c	n	k	...	s	w	w	p	w	o	p	k	
1	e	x	s	y	t	a	f	c	b	k	...	s	w	w	p	w	o	p	n	
2	e	b	s	w	t	l	f	c	b	n	...	s	w	w	p	w	o	p	n	
3	p	x	y	w	t	p	f	c	n	n	...	s	w	w	p	w	o	p	k	
4	e	x	s	g	f	n	f	w	b	k	...	s	w	w	p	w	o	e	n	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
8119	e	k	s	n	f	n	a	c	b	y	...	s	o	o	p	o	o	p	b	
8120	e	x	s	n	f	n	a	c	b	y	...	s	o	o	p	n	o	p	b	
8121	e	f	s	n	f	n	a	c	b	n	...	s	o	o	p	o	o	p	b	
8122	p	k	y	n	f	y	f	c	n	b	...	k	w	w	p	w	o	e	w	
8123	e	x	s	n	f	n	a	c	b	y	...	s	o	o	p	o	o	p	o	

8124 rows × 23 columns

General: There are 8123 observations of mushrooms with no null value(s) present in each column

Feature Selection

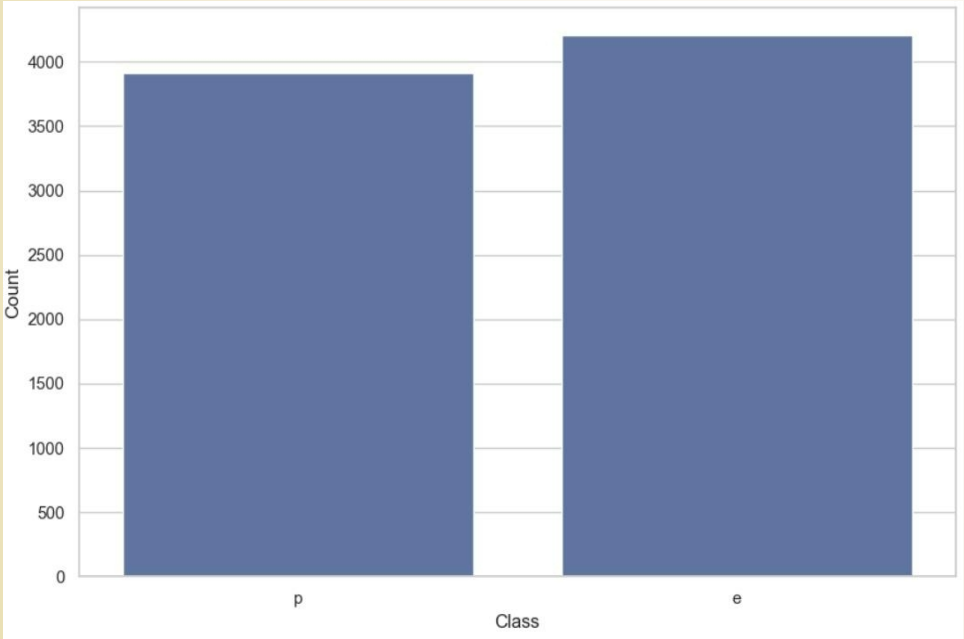
Out of 22 features, 9 features that are most *visually apparent* and *relevant* are selected

- Cap Features: Shape and surface
- Gill Features: Attachment, spacing
- Stalk Feature: Shape
- Ring Features: Number and type
- Bruises: Presence or absence
- Habitat: Type of surrounding environment where the mushroom is found

Attribute Description

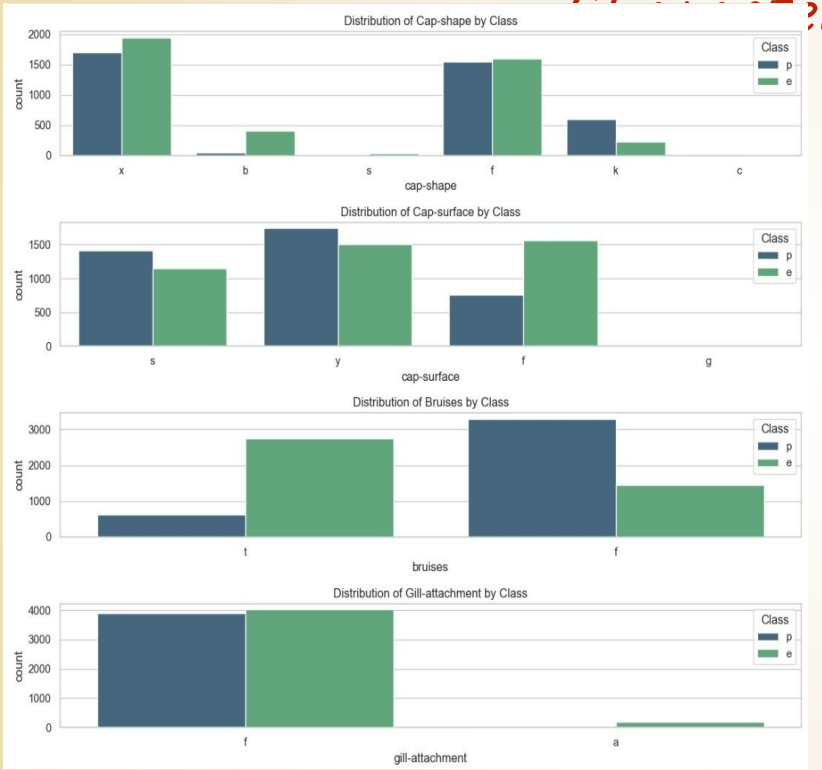
Attribute	Description
classes	edible = e, poisonous = p
cap-shape	bell = b, conical = c, convex = x, flat = f, knobbed = k, sunken = s
cap-surface	fibrous = f, grooves = g, scaly = y, smooth = s
bruises	bruises = t, no = f
gill-attachment	attached = a, descending = d, free = f, notched = n
gill-spacing	close = c, crowded = w, distant = d
stalk-shape	enlarging = e, tapering = t
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
habitat	grasses = g, leaves = l, meadows = m, paths = p, urban = u, waste = w, woods = d

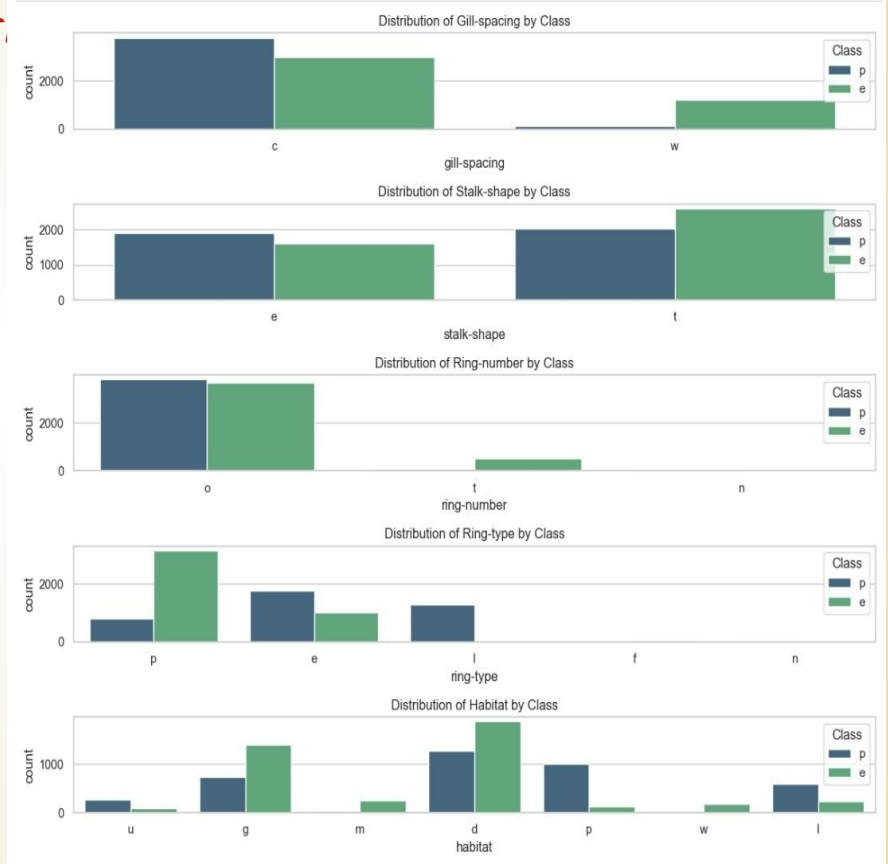
Mushroom Data: Proportion of Edible vs. Poisonous Mushrooms



- 4,208 edible (e) and 3,916 poisonous (p) mushrooms
- Fairly balanced dataset
- Beneficial for modeling since it helps prevent a model from being biased towards the more frequent class

Mushroom Dataset: Exploring Mushroom Characteristics



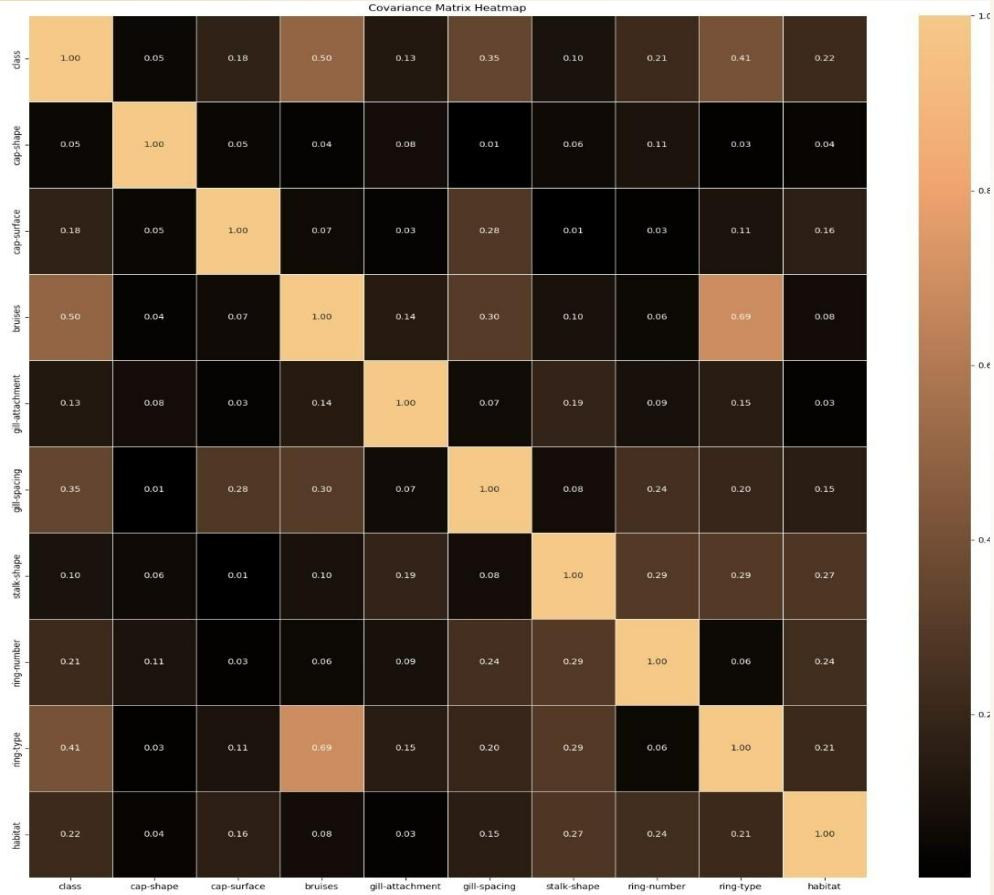


Mushroom Dataset: Exploring Mushroom Characteristics (Cont.)

- Cap Characteristics: Variability in cap shape and surface suggests these attributes could be significant in differentiating between classes, with certain types possibly leaning towards one class over the other
- Bruises: The occurrence of bruises on mushrooms appears to be a prominent feature that could be used to distinguish between edible and poisonous classes, with one class showing a higher tendency towards bruising
- Gill Attachment: The type of gill attachment, especially the free type, is indicative of the mushroom's class, highlighting its potential as a strong predictor in classification tasks
- Gill Spacing: The space is distinctly distributed between the classes, suggesting it is informative for predicting whether a mushroom is edible or poisonous
- Stalk Features: The shape of the stalk show differences between classes
- Ring Characteristics: The number and type of rings offer valuable clues; no rings or two rings suggest edibility, whereas one ring may suggest toxicity. Moreover, pendant rings are commonly found on edible mushrooms
- Habitat: The habitat of a mushroom is a significant factor, with urban and wooded areas more likely to harbor poisonous varieties, and grassy areas and meadows more likely to have edible ones

Takeaway: The distinct distribution of cap characteristics, bruises, gill attachment and spacing, stalk features, ring types, and habitat environments provides critical insights, making these attributes *highly effective for distinguishing between edible and poisonous mushrooms in classification tasks*

Mushroom Dataset: Covariance Matrix Heatmap



Class Correlation:

- **Key predictors: *bruises* have a notable correlation with class**

Multicollinearity Concerns:

- ***Ring-type* and *bruises* have high correlation**
  - **May impact models that assume feature independence**

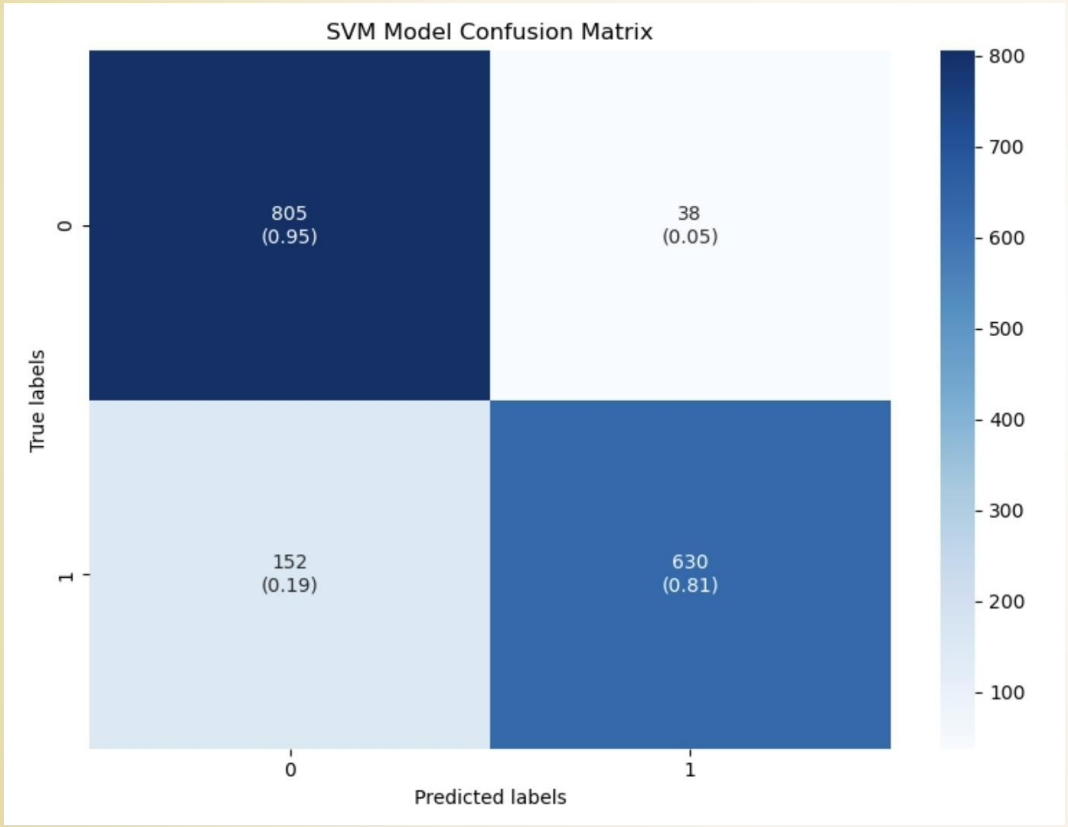
04

Model

Training Steps

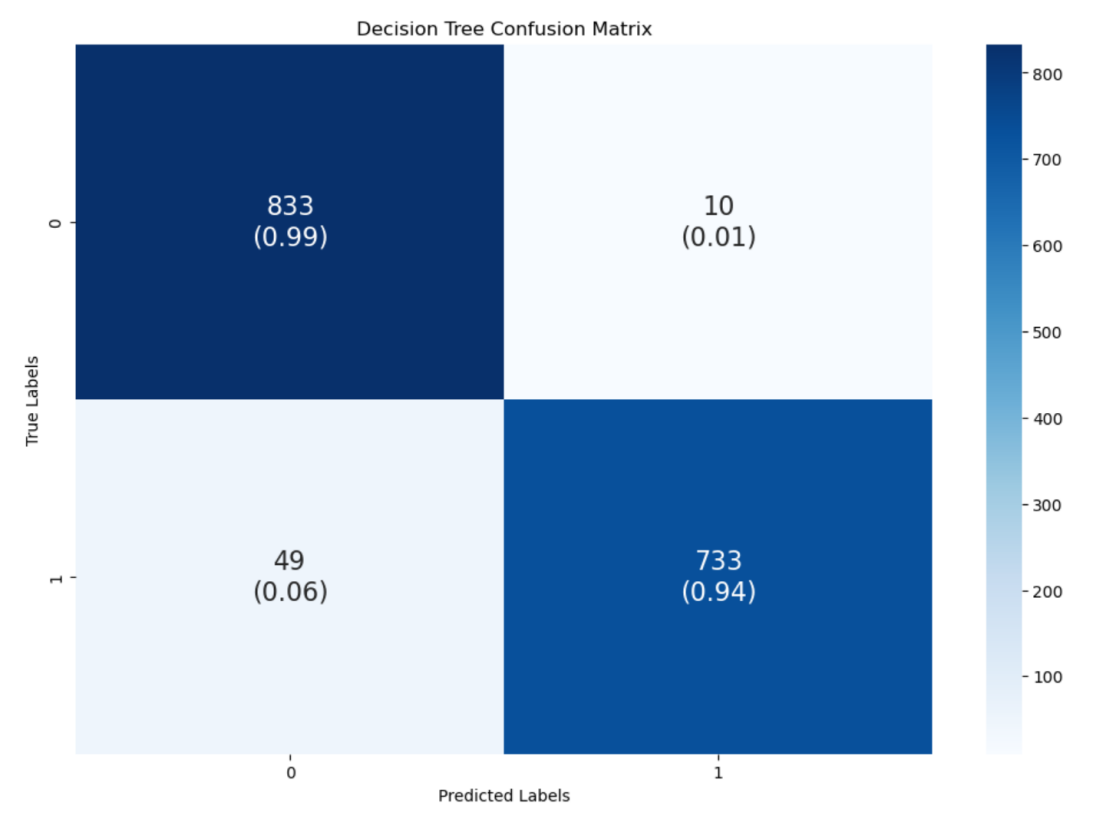
- Split into training set and testing set
  - Training set: 0.8
  - Testing set: 0.2
- Encode and standardize categorical Variable
- Model candidates: Decision Tree, Random Forest, Boost Tree, SVM
- Hyperparameter Tuning
  - Scorer: Negative Predicted Value
    - We want to minimize the probability of being poisonous given predicted as edible
  - Decision Tree Parameters
    - Criterion: gini, entropy
    - Maximum depth: none, 1, 2, 3, 4, 5
  - Random Forest Parameter:
    - Number of Estimators: 10, 50, 100, 150
    - Maximum depth : none, 1, 2, 3, 4, 5
  - Boost Tree Parameters:
    - Number of Estimators: 10, 50, 100
    - Maximum depth: none, 1, 2, 3, 4, 5
    - Learning Rate: 0.01, 0.1, 0.2, 0.5
  - SVM Parameters
    - Cost: 0.1, 1
    - Gamma: 1, 10

Best Support Vector Machine



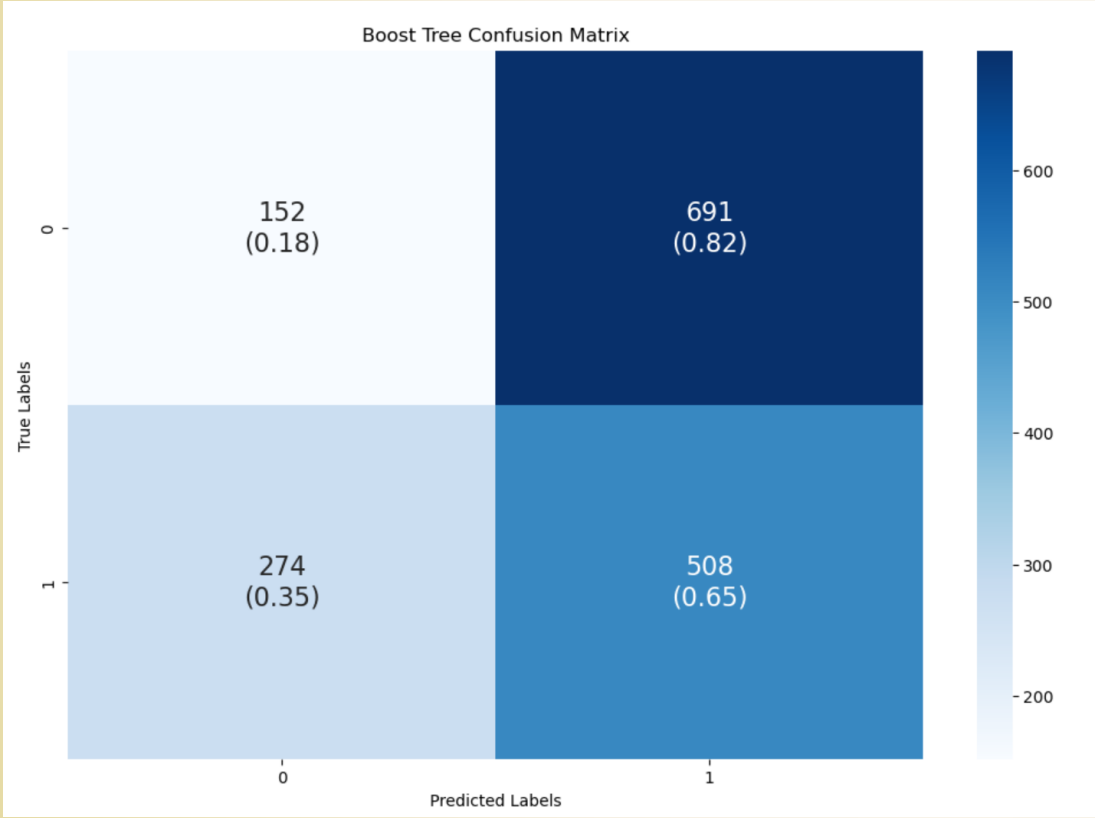
- Best parameters
  - Regularization: 1
  - Gamma: 1
  - Kernel: Linear
- Accuracy = 0.90
- Negative Predicted Value = 0.806

Best Decision Tree



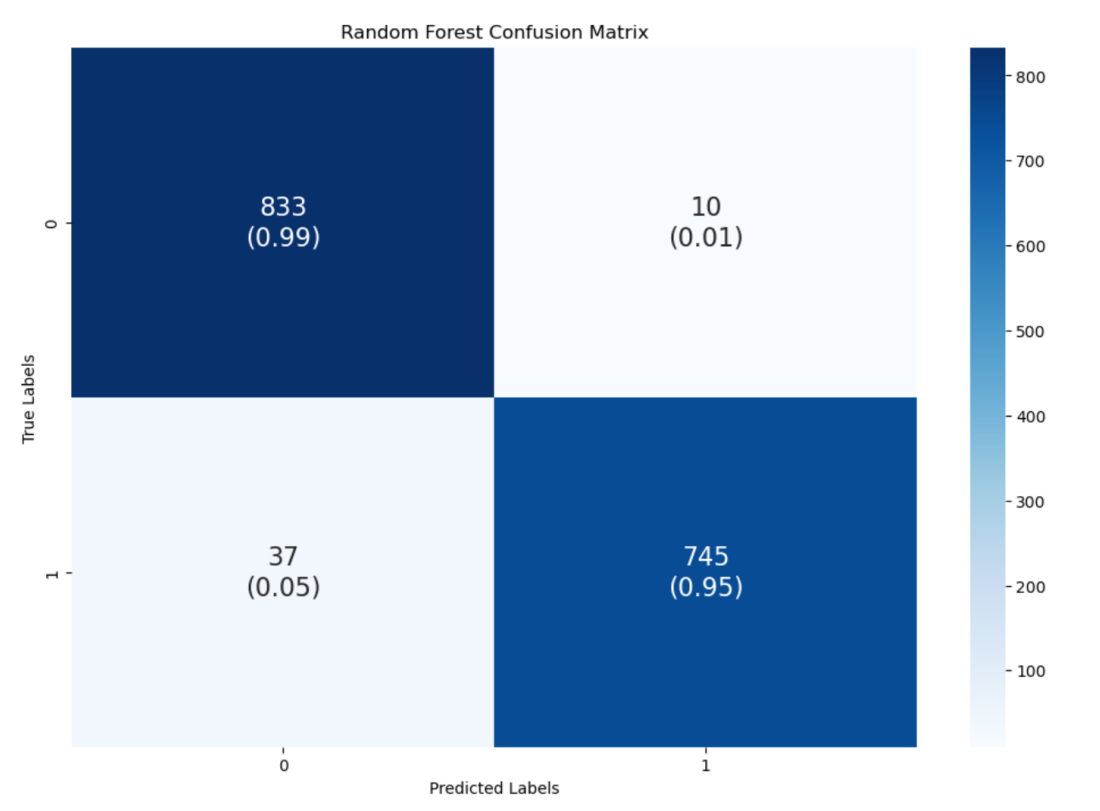
- Best parameters
  - Criterion: gini
  - Maximum depth:5
- Accuracy = 0.95
- Negative Predicted Value = 0.944

Best Boost Tree



- Best parameters
  - Number of Estimators: 100
  - Maximum depth: 5
  - Learning Rate: 0.5
- Accuracy = 0.41
- Negative Predicted Value = 0.357

Recommend Random Forest



- Best parameters
  - Number of Estimators: 50
  - Maximum depth: 5
- Accuracy = 0.97
- Negative Predicted Value = 0.957

05

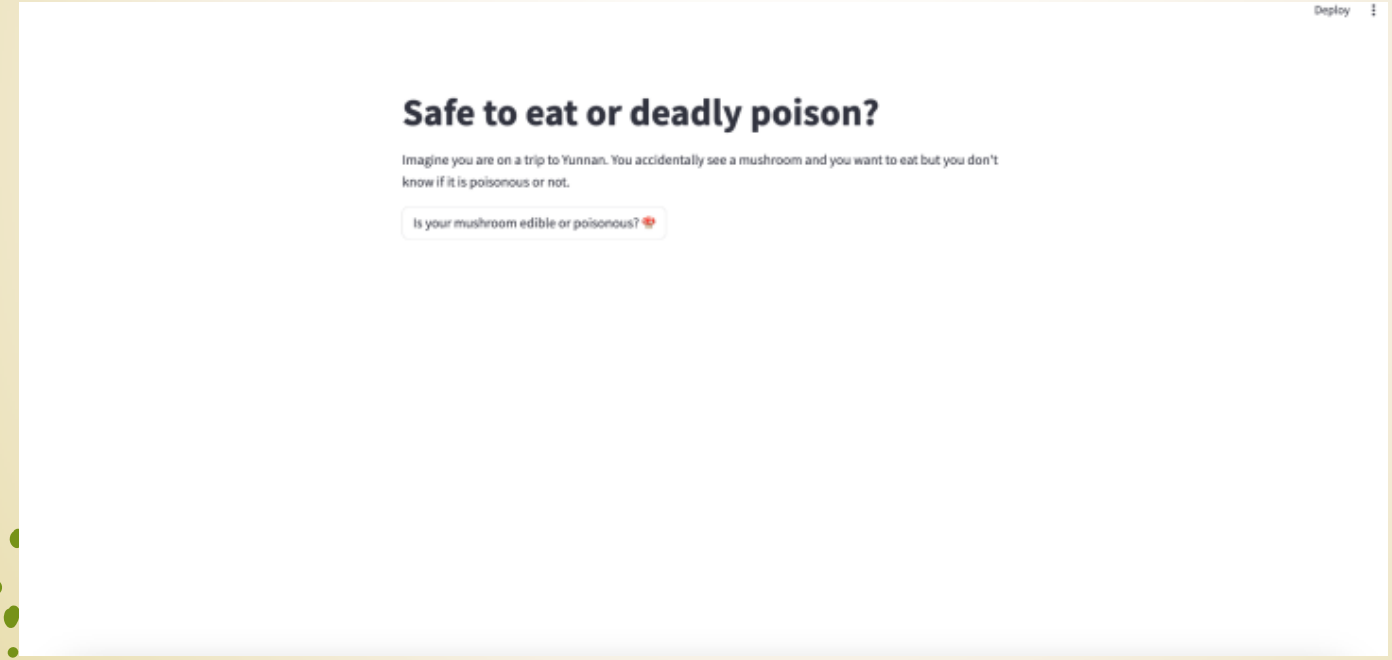
Application

Design Idea

- When people happen to see a mushroom and wonder if it is edible, they can use our application to recognize toxicity
- The application provides instant feedback on the toxicity of the mushroom based on the models we trained
- The application mentions the high accuracy rate (97%) of the recognition system in identifying toxic mushrooms
- The interface is easy to read and easy to understand

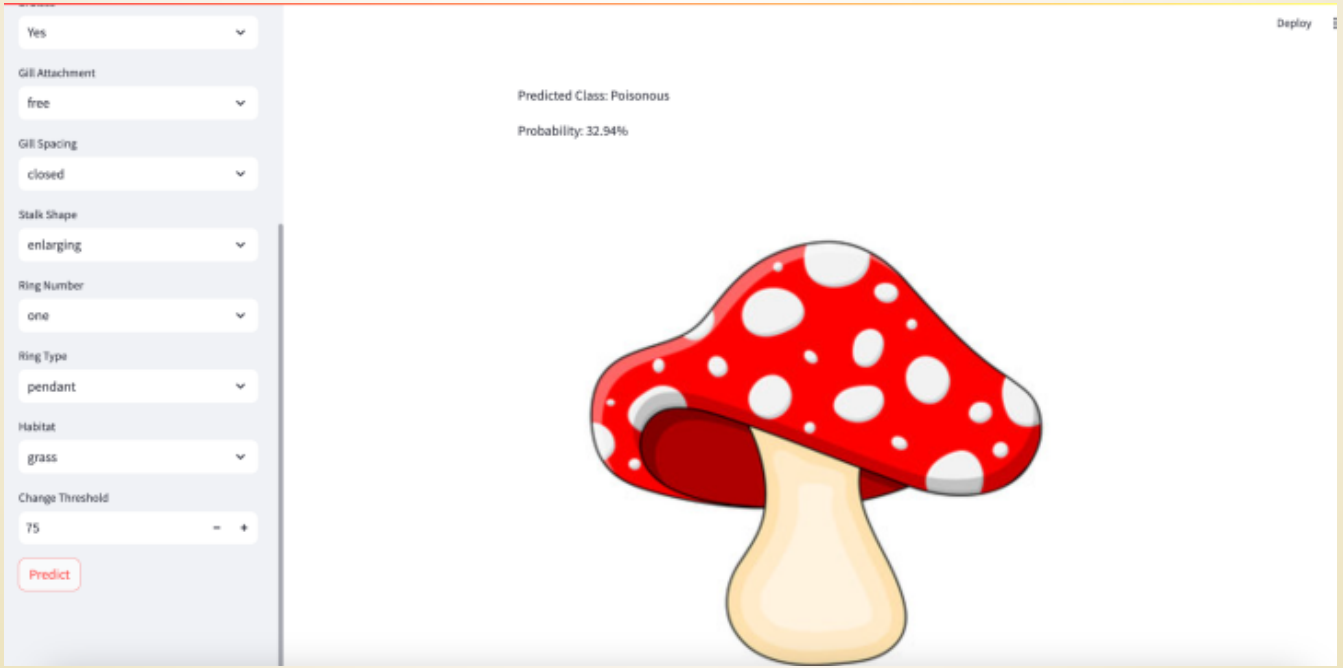
Streamlit Application

· A streamlit Application was made to help people recognize the toxicity of mushrooms and help reduce accidents



Streamlit Application

· A streamlit Application was made to help people recognize the toxicity of mushrooms and help reduce accidents



Conclusion

06

Our Methodology Summary

Model

Based on the test of Linear Regression, Support Vector Machine(SVM), Random Forest, Decision Tree and XGBoost, we use **Random Forest** as our final solution

Application

StreamLit
(We focus on user experience)

Our Preliminary Plan

1. Test out the apps we design and seek feedback from users

2. Seeking more potential models with higher performance

3. Further optimize our StreamLit app and design the UI/UX

THANK YOU !

-SINO DRAGON