Fungi or Foe?

A Statistical Exploration of Classification Problems with Mushroom Edibility

By: Jenna Sallin and Madison Owen

Table of Contents

PART 01

Introduction

Mushroom Facts

It is estimated that there are over 5000 species of mushrooms worldwide

Mushroom poisonings are estimated to cause over 10,000 illnesses and 100 deaths annually

Of these 5000 species, only 20-25% have been named, with 3% identified as poisonous

In North America alone, there is estimated to be over 250 poisonous species of mushrooms

The Problem

Current Ways of Identifying Mushrooms:

- Joining a foraging club or group, and learning from an expert
- Memorizing the characteristics of any mushroom you could encounter in your ecosystem
- Inspecting the mushroom under a microscope (only definitive way)

All of these require extensive knowledge, funding, and the willingness to take a risk on the edibility of your mushroom

- As of today, there is no app that can positively identify a mushroom through technology

Our Solution

Work to create a machine-learning model to aid in mushroom classification. Our model would allow foragers to input characteristics of mushrooms they might encounter, and would output the edibility of the given mushroom. This model would be easily interpretable to the general public and would provide ease of use to those without a strong mathematical background

Research Questions

01 Which types of machine learning models are the most accurate on our data?

02 What variables are the most significant in classification for each model?

03 Are parametric or non-parametric models better for our data?

04 How do our models perform in a real-life scenario? Are they reliable?

Goals for Our Model

Easily interpretable to the public

Simple to use

Highly accurate

Applicable to real-life scenarios

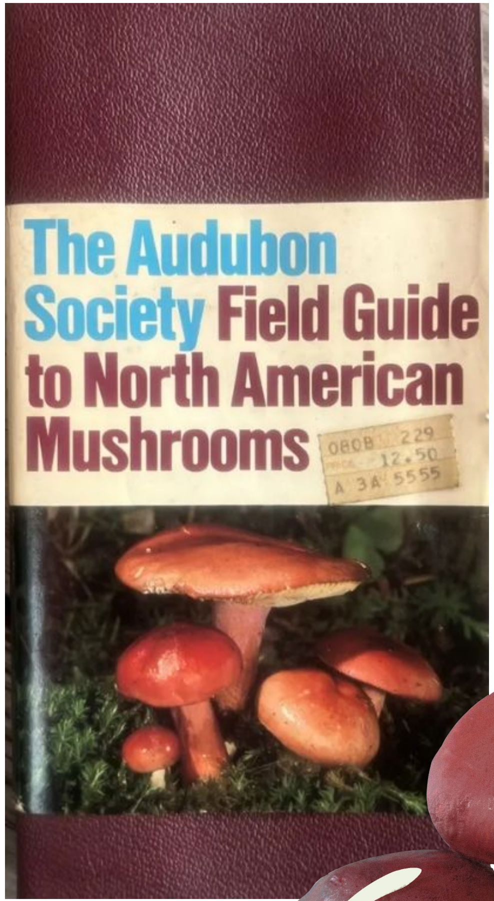
PART 02

Description of the Data

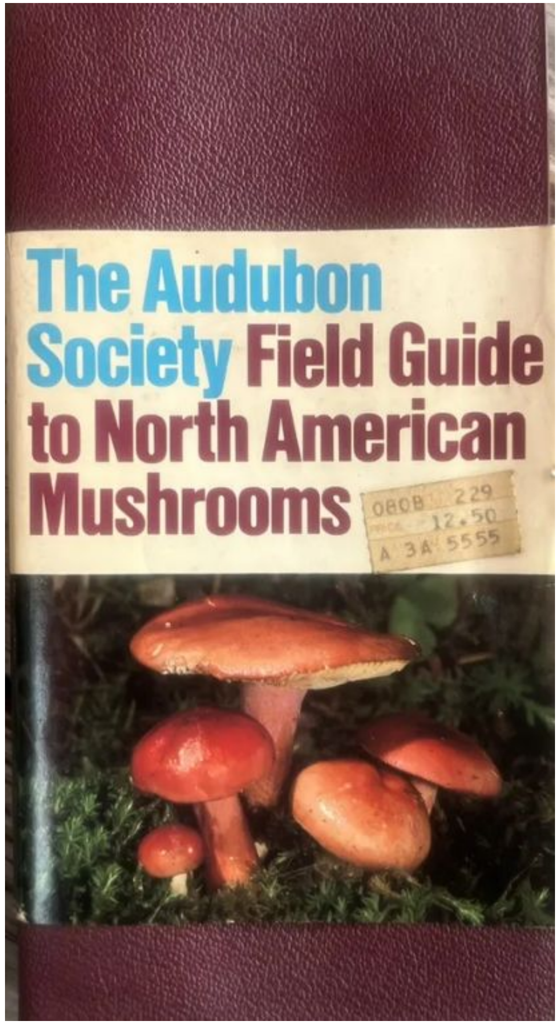
About the Dataset

Mushroom Classification:

- Dataset acquired from Kaggle
  - Originally contributed to the UCI Machine Learning Repository on April 27th, 1987
- The 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)



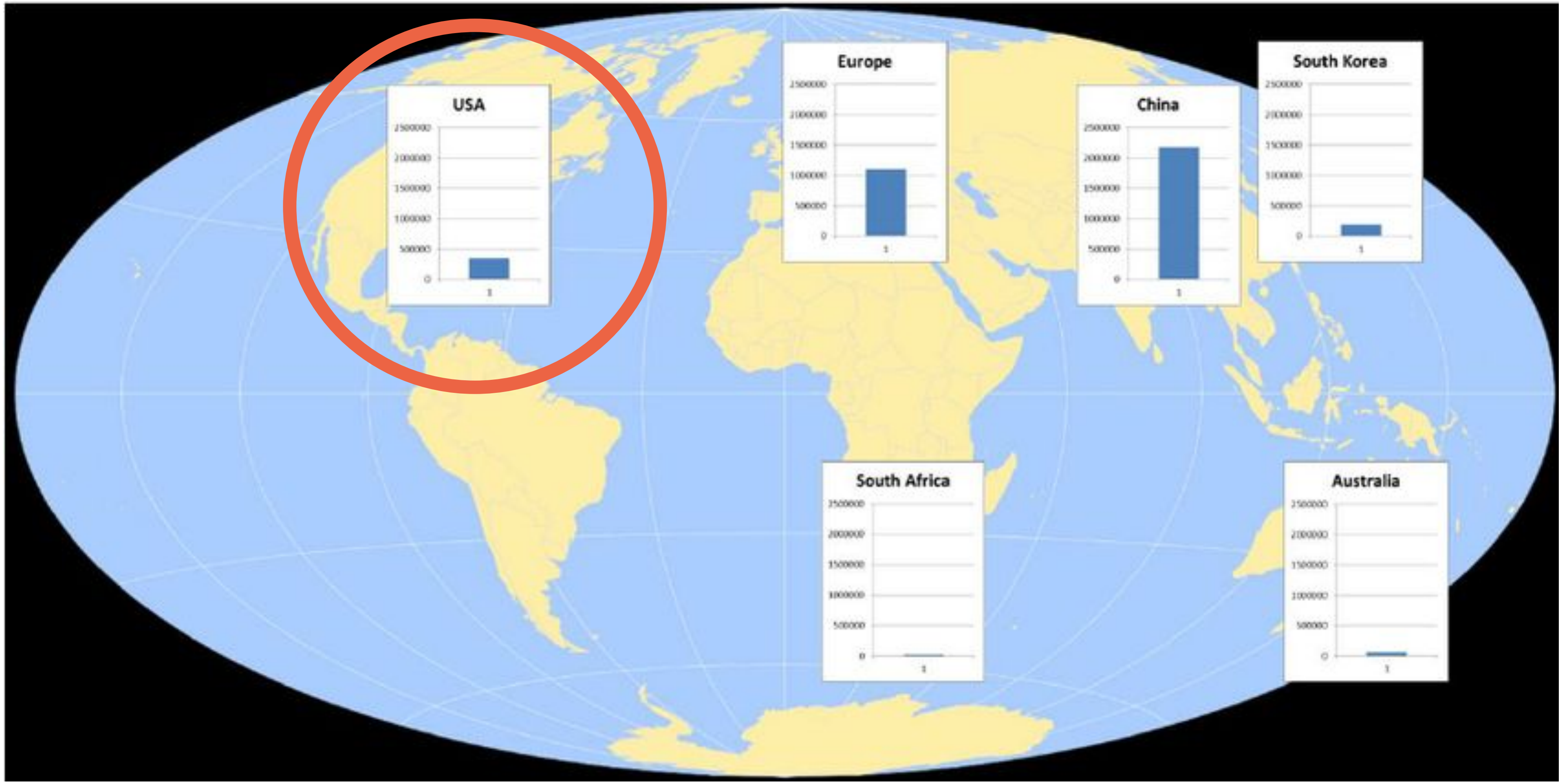
The Audubon Society Field Guide to North American Mushrooms (1981)



- Details over 700 species of mushrooms, grouping mushrooms by color and shape
- Includes a section on cooking and eating wild mushrooms,
- Each species includes a detailed physical description, information on edibility, season, habitat, range, look-alikes, alternative names, and facts on edible and poisonous species, uses, and folklore

Agaricus and Lepiota Family Habitats

Global distribution of *Agaricus bisporus* production



Agaricus and Lepiota Family Facts

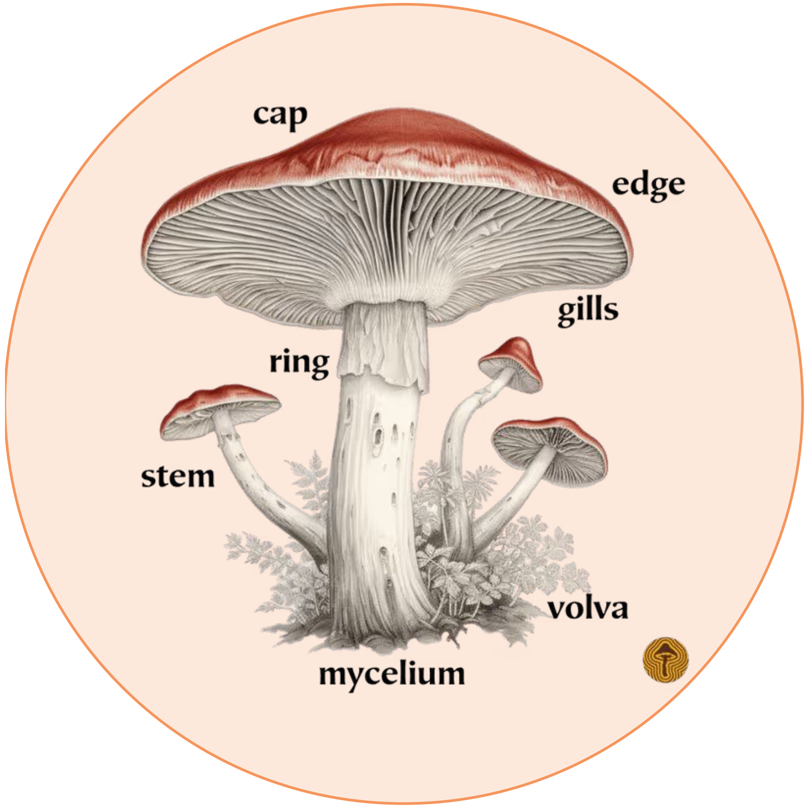
Agaricus Bisporus account for nearly 90% of the mushroom production in the United States!

Over 400 species in these families are recognized worldwide!

They are also known as button mushrooms!

Important Common Attributes of Mushrooms

Attributes	Description
cap shape	The shape of the upper part of the mushroom (bell, conical, etc.)
cap color	Color of the upper part of the mushroom (brown, green, red, etc.)
cap surface	Texture of the upper part (grooves, smooth, etc.)
bruises	Discolorations or marks that appear on the flesh of a mushroom
odor	How the mushroom smells (musty, pungent, etc.)
gill spacing	Distance between the gills on the underside of a mushroom's cap
gill size	The size of the part under the cap (i.e. the gill)
stalk shape	The shape of the bottom most part (enlarging, tapering)
stalk root	Shape of the root of the mushroom (bulbous, club, etc.)
veil type	Form of a mushroom's protective covering over its gills or spore surface
ring type	Texture of the mushroom’s rings (cobwebby, flaring, etc.)
population	Number of mushrooms grown from one root
habitat	The place where the mushroom is grown (meadows, woods, etc.)



Our Dataset Includes

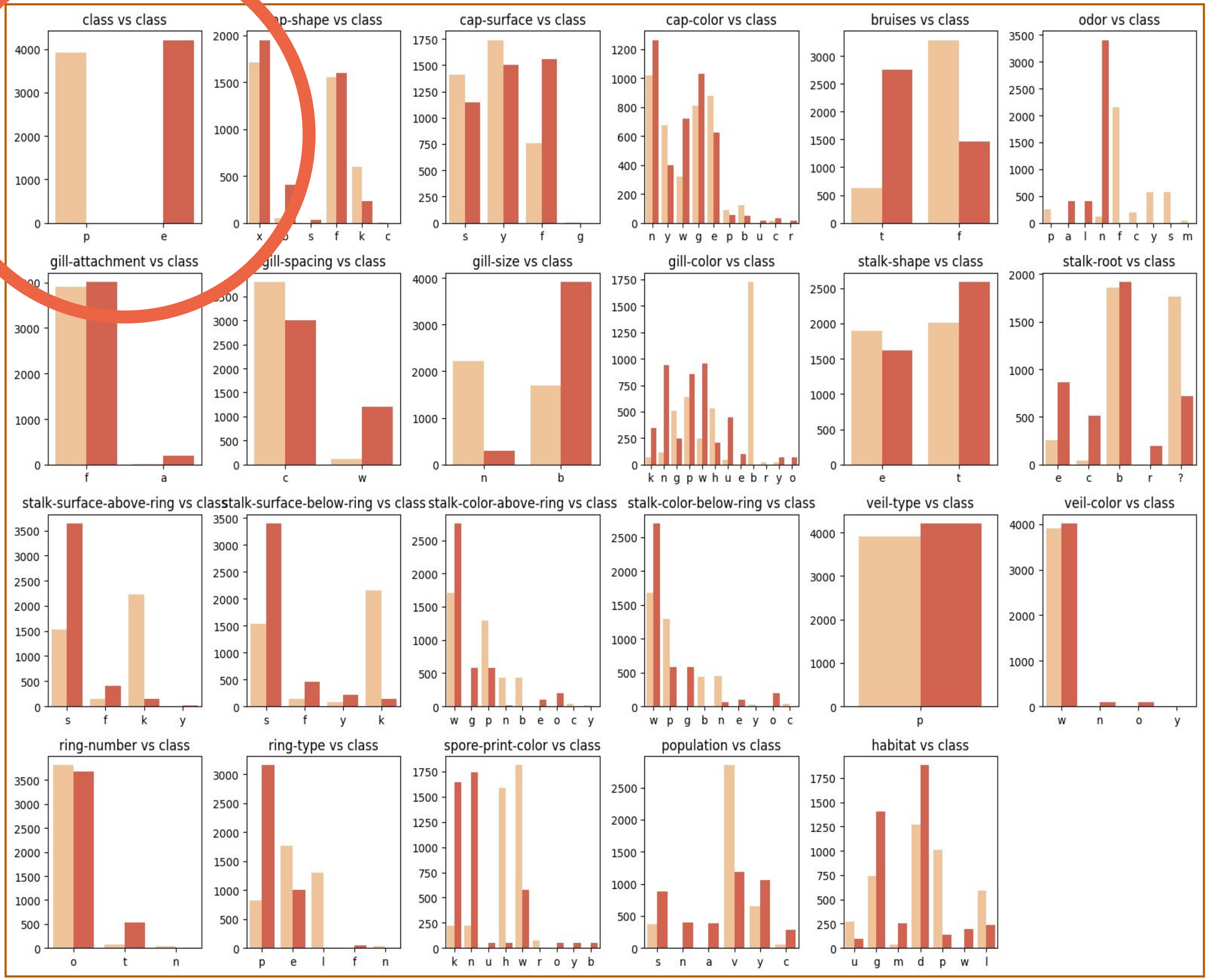
- **8,124 samples**
  - 4,208 edible and 3,916 poisonous
- **23 different categorical variables**
  - We will be using “Class” to indicate whether a sample is poisonous or edible

PART 03

Data Visualization

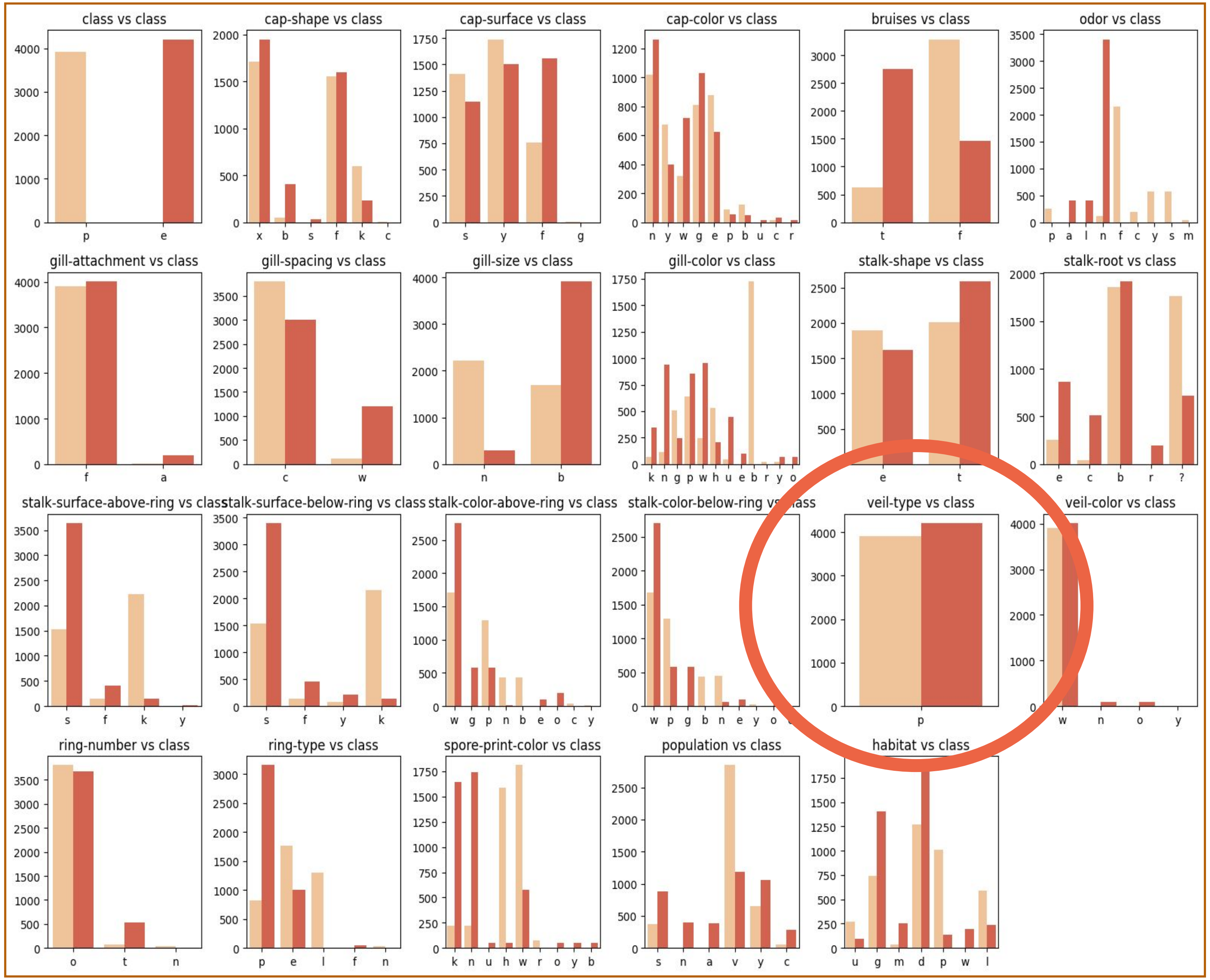
Histograms of our Variables

Our data is fairly balanced for the “Class” variable, which we will be using as our response. We see nearly half of the samples are poisonous while the other half is edible



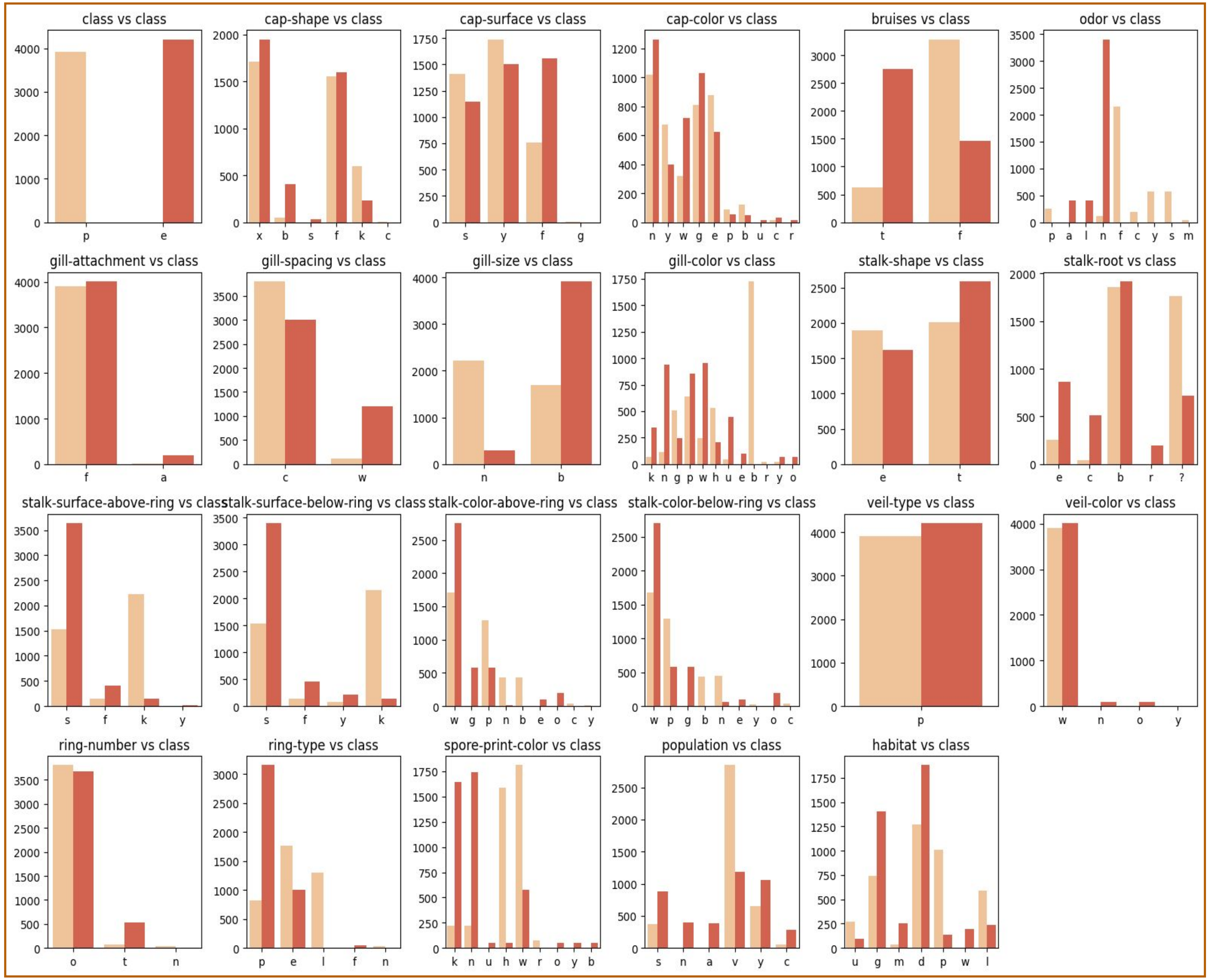
Histograms of our Variables

We observe that all 8,124 samples have the same veil type (p) so we drop this variable from the model

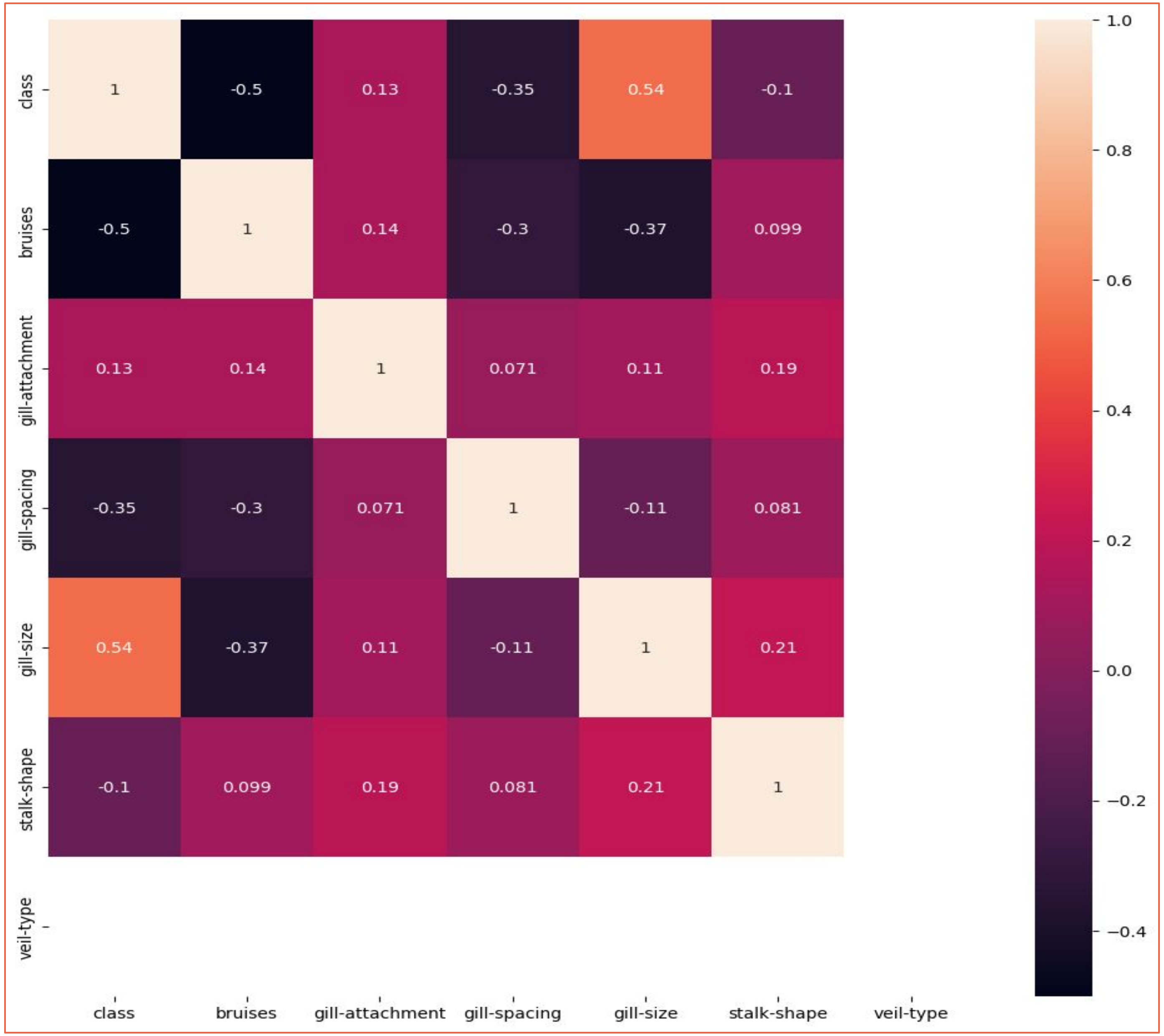


Histograms of our Variables

For the remaining variables, the number of edible and poisonous mushrooms is comparable, with no clear attribute that poisonous mushrooms have over edible mushrooms, and visa versa



Correlation Matrix



PART 04

Data Cleaning

Creating Dummies- Part 1

For our variables with two classes, we created single dummy variables:

$$\text{Class} = \begin{cases} 1 \text{ if poisonous} \\ 0 \text{ if edible} \end{cases}$$

$$\text{Gill Attachment} = \begin{cases} 1 \text{ if free} \\ 0 \text{ if attached} \end{cases}$$

$$\text{Bruises} = \begin{cases} 1 \text{ if has bruises} \\ 0 \text{ if not} \end{cases}$$

$$\text{Stalk Shape} = \begin{cases} 1 \text{ if enlarging} \\ 0 \text{ if tapering} \end{cases}$$

$$\text{Gill Spacing} = \begin{cases} 1 \text{ if crowded} \\ 0 \text{ if close} \end{cases}$$

$$\text{Gill Size} = \begin{cases} 1 \text{ if narrow} \\ 0 \text{ if broad} \end{cases}$$

Creating Dummies- Part 2

For our remaining variables with more than two classes (say k classes):

We created k-1 dummy variables for each of these categorical variables

- 1 is assigned to the attribute that the mushroom expresses
- 0 is assigned to the remaining attributes that the mushroom does not express

	class	bruises	gill-attachment	gill-spacing	gill-size	stalk-shape	veil-type	cap-shape_c	cap-shape_f	cap-shape_k	...
0	1	1	1	0	1	0	0	0	0	0	...
1	0	1	1	0	0	0	0	0	0	0	...
2	0	1	1	0	0	0	0	0	0	0	...
3	1	1	1	0	1	0	0	0	0	0	...
4	0	0	1	1	0	1	0	0	0	0	...

Dependent and Independent Variables

Dependent Variable

$$\text{Class} = \begin{cases} 1 \text{ if poisonous} \\ 0 \text{ if edible} \end{cases}$$

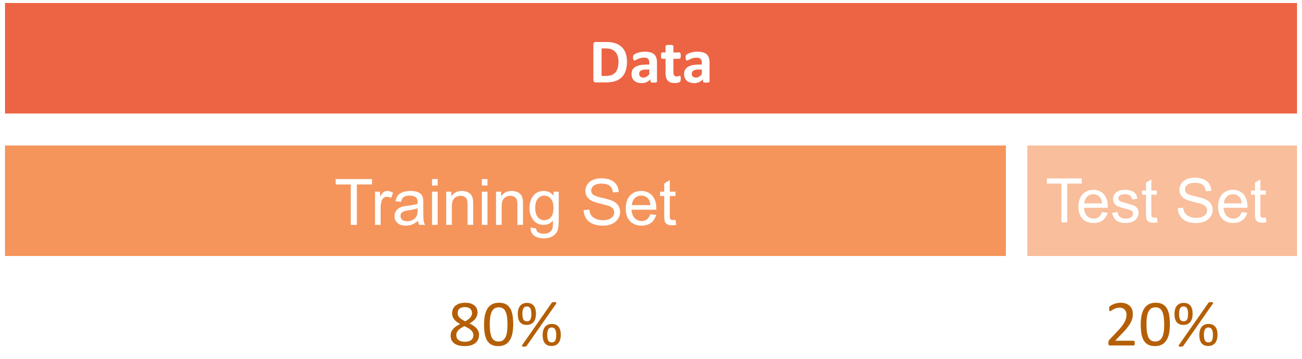
Independent Variables

cap surface, cap shape, cap color, bruises, odor, gill attachment, gill spacing, gill size, gill color, stalk shape, stalk root, stalk surface above ring, stalk surface below ring, stalk color above ring, stalk color below ring, veil color, ring number, ring type, spore print color, population, & habitat

Splitting Into Training and Test Sets

Training set- 80% of our data

Test set- 20% of our data




```
x_train, x_test, y_train,y_test = train_test_split(x, y, test_size = 0.2, stratify = y, random_state = 42)
```

PART 05

Model Creation

Our 8 Classification Models

- Logistic Regression
- Decision Tree
- Boosting
- Naive-Bayes
- K-Nearest Neighbors (KNN)
- Random Forest
- Support Vector Machine Classification (SVM)
- Linear Discriminant Analysis (LDA)

Model 1:
Logistic Regression

About Logistic Regression

Logistic Regression is:

- **A PARAMETRIC MODEL**
  - Assumes a linear relationship between the predictors and the log-odds of the response, independence of observations, and absence of multicollinearity, outliers, perfect separation, and endogeneity.
- **USED FOR BINARY CLASSIFICATION**

We Chose Logistic Regression Because of its:

- **INTERPRETABILITY**
- **SUITABILITY IN REAL LIFE SCENARIOS**
- **ABILITY TO SERVE AS A BASELINE TO COMPARE OTHER MODELS**

Formulas for Logistic Regression

$$\mathrm{logit}(\mathbb{P}(Y = 1|\boldsymbol{X} = \boldsymbol{x})) = \log\left(\frac{p(\boldsymbol{x})}{1 - p(\boldsymbol{x})}\right)$$

$$\mathbb{P}(Y = 1|\boldsymbol{X} = \boldsymbol{x}) = \frac{e^z}{1 + e^z}$$

We choose the class j, either 0 or 1 in this case, for which the probability above is at its maximum

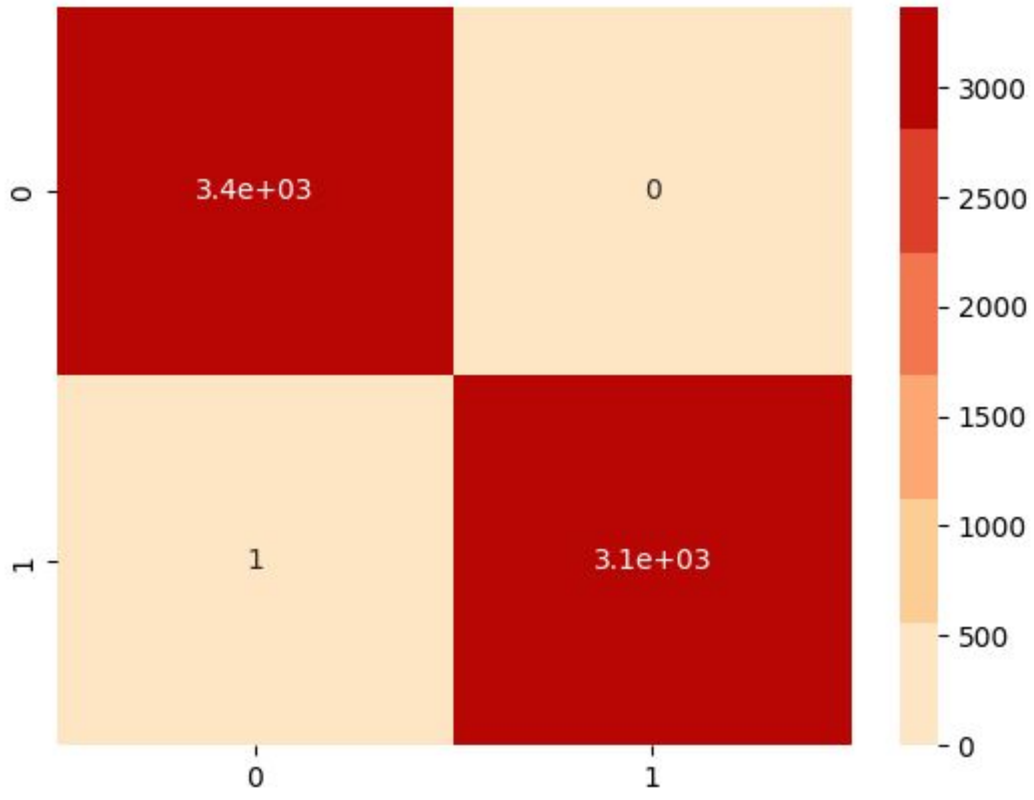
Confusion Matrix

A confusion matrix shows a comparison of the predicted classes to the actual classes of a set of data

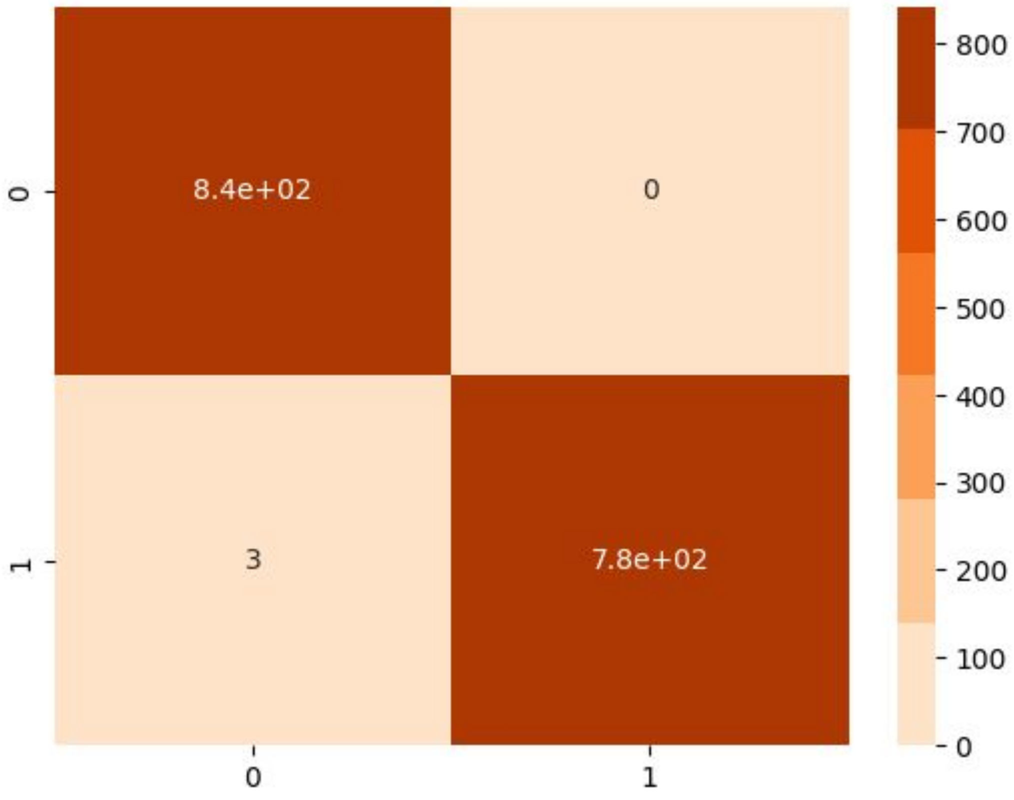
	Predicted Edible	Predicted Poisonous
Actually Edible	True Edible	False Poisonous
Actually Poisonous	False Edible	True Poisonous

Confusion Matrix for Logistic Regression

Training:



Test:



Accuracies

Training Accuracy = 0.9998461301738729

Test Accuracy = 0.9981538461538462

Model 2:
K-Nearest Neighbors (KNN)

About K-Nearest Neighbors (KNN)

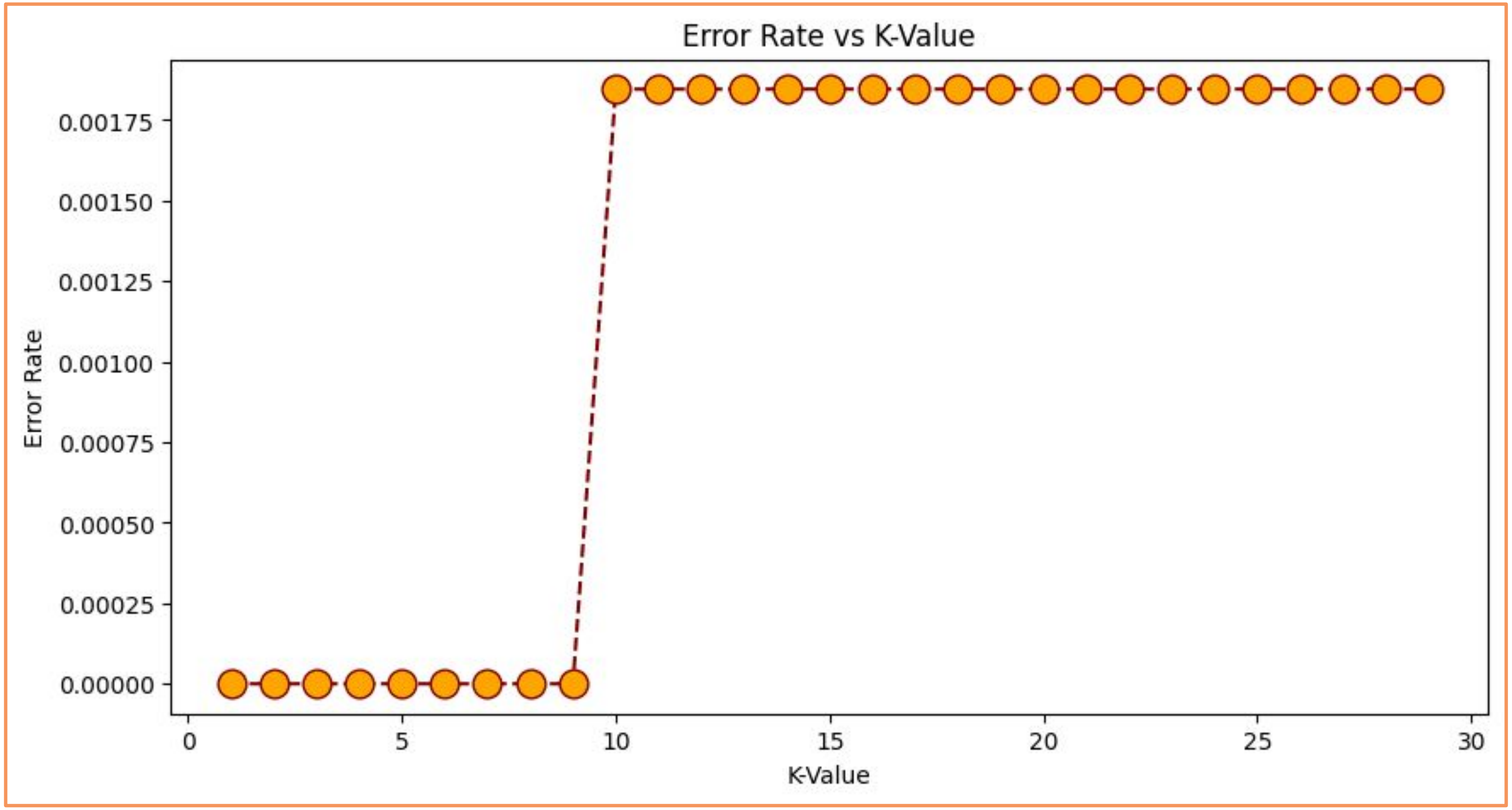
KNN is:

- **A NON-PARAMETRIC MACHINE LEARNING ALGORITHM**
- **USED FOR CLASSIFICATION AND REGRESSION TASKS**
  - Done by relying on proximity to determine the class of a data point based upon its neighbors

We Chose KNN Because of its:

- **SIMPLICITY**
- **ADAPTABILITY TO COMPLEX DECISION BOUNDARIES**
- **SMALLER SENSITIVITY TO OUTLIERS**
- **VERSATILITY**

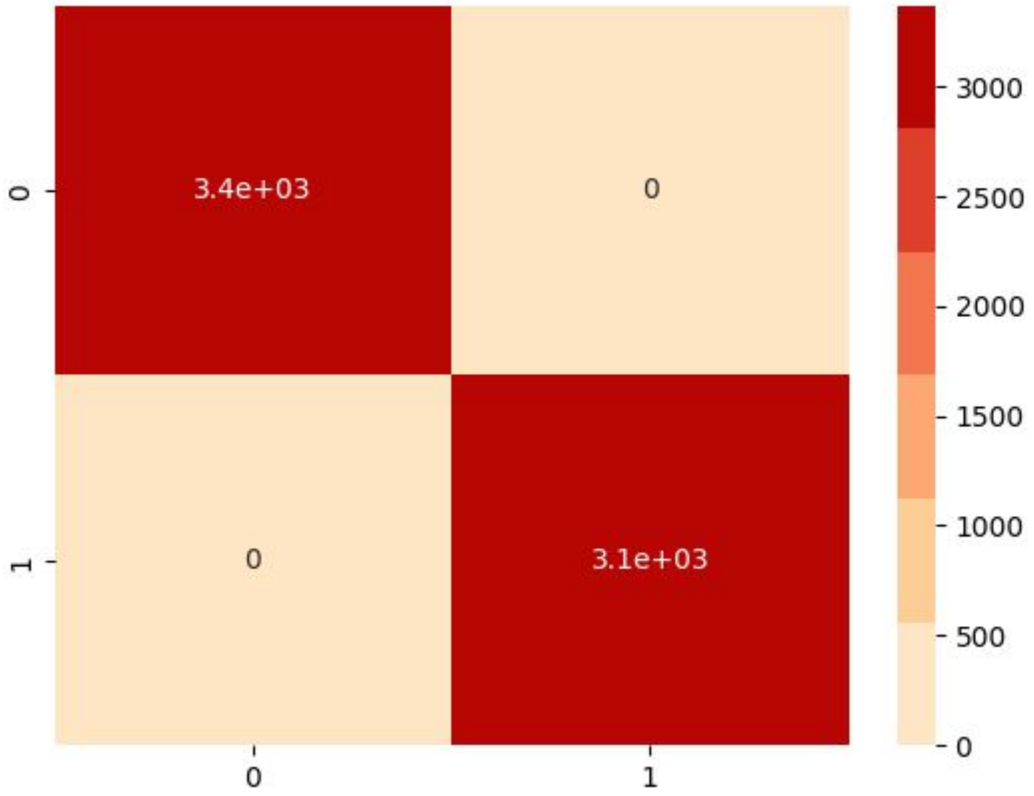
Picking Our Number of Neighbors



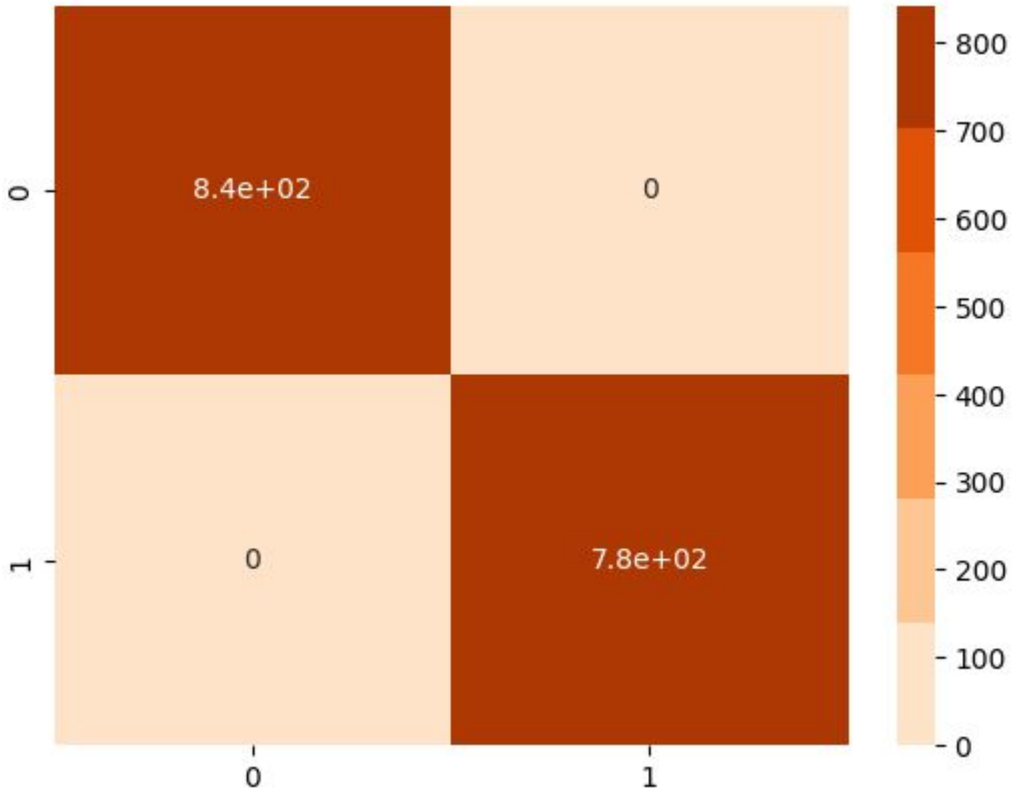
In order to minimize our error, we choose K=3 neighbors

Confusion Matrix for KNN

Training:



Test:



Accuracies

Training Accuracy = 1.0

Test Accuracy = 1.0

Model 3: Decision Tree

About Decision Trees

Decision Trees are:

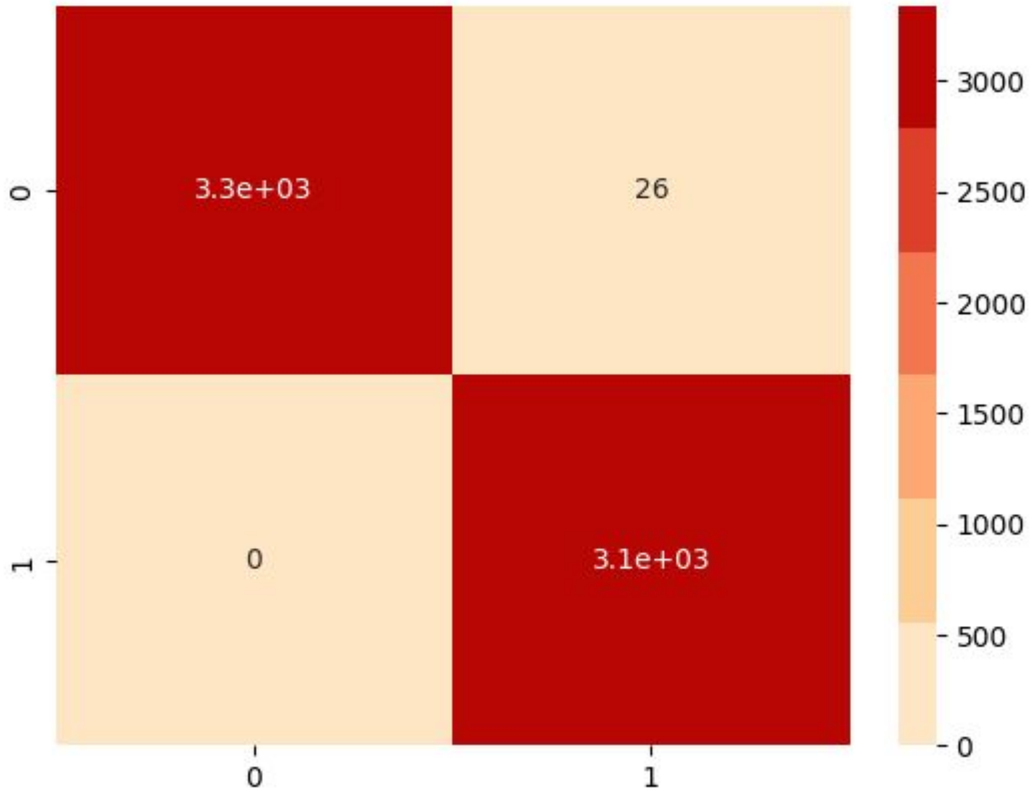
- **NON-PARAMETRIC TREE-LIKE MODELS FOR CLASSIFICATION**
  - each internal node represents a decision based upon a specific feature, leading to leaf nodes representing the final outcome

We Chose Decision Trees Because of Their:

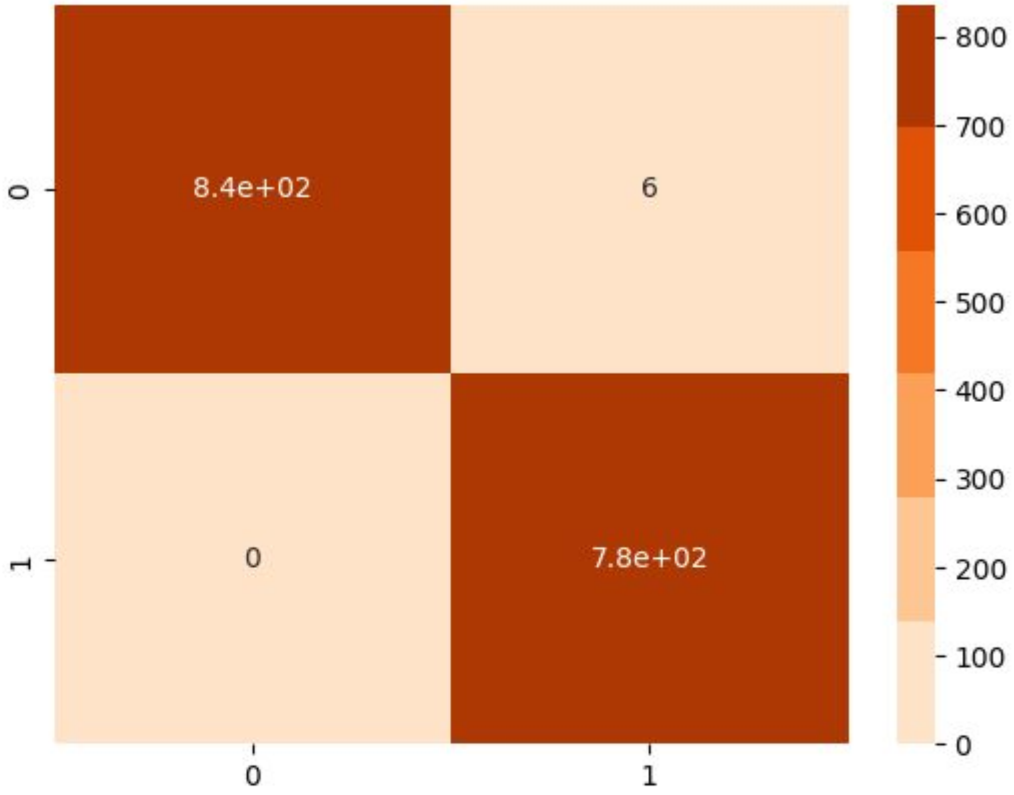
- **ABILITY TO CAPTURE COMPLEX RELATIONSHIPS IN OUR DATA**
- **INTERPRETABILITY**
- **EASE OF USE**
- **APPLICABILITY TO REAL-WORLD MUSHROOM CLASSIFICATION**

Confusion Matrix for Decision Tree

Training:



Test:



Accuracies

Training Accuracy = 0.9959993845206955

Test Accuracy = 0.9963076923076923

Model 4:
Random Forest

About Random Forest

Random Forest is:

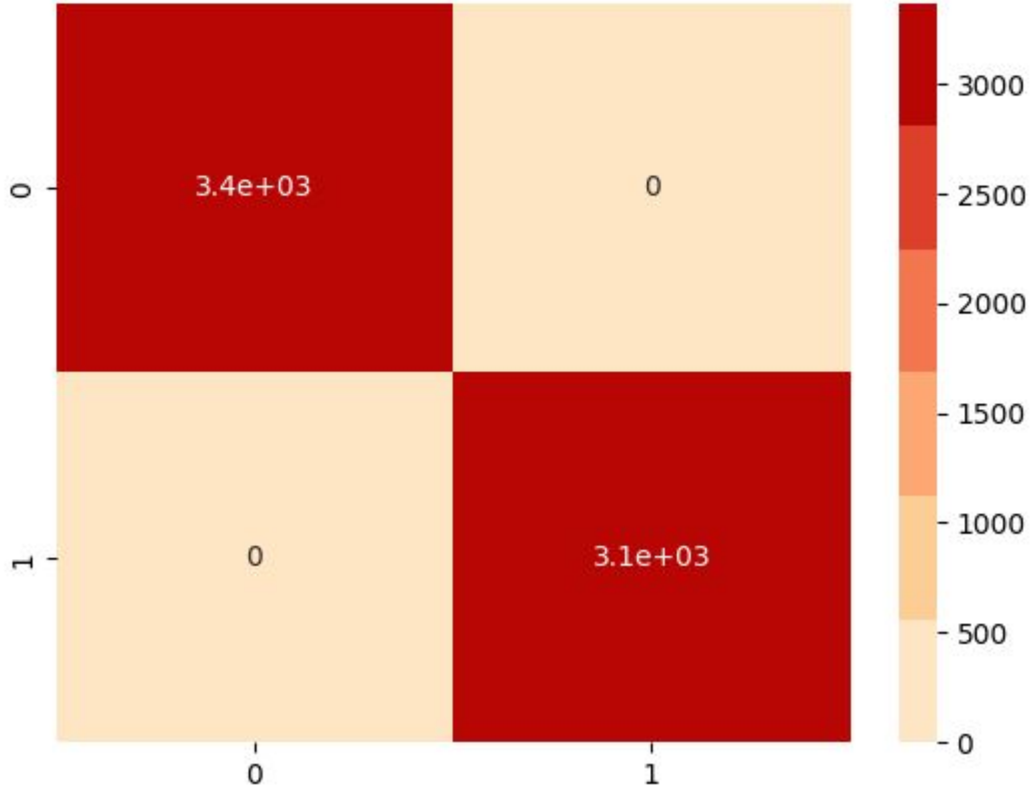
- **A NON-PARAMETRIC LEARNING METHOD**
  - Constructs multiple decision trees during training and outputs the mode of the classes for classification
- **USED FOR CLASSIFICATION PROBLEMS**

We Chose Random Forest Because of its:

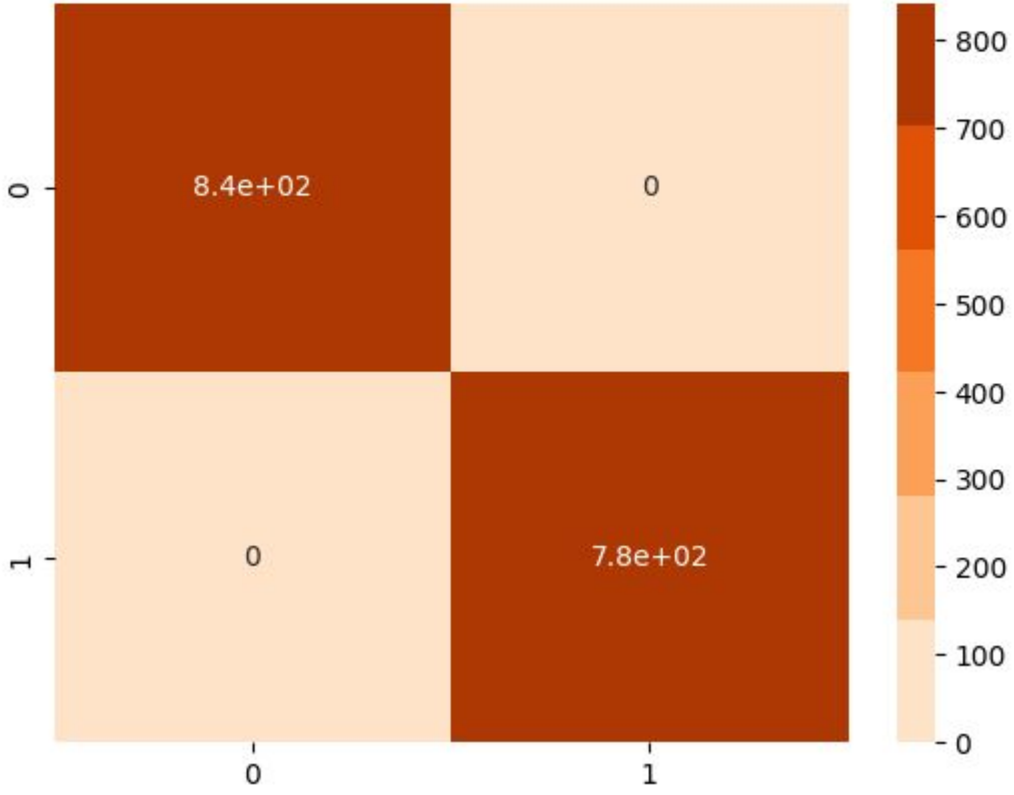
- **ABILITY TO HANDLE COMPLEX RELATIONSHIPS IN OUR DATA**
- **ABILITY TO REDUCE OVERFITTING**
- **CAPABILITY OF PROVIDING ROBUST PREDICTIONS**

Confusion Matrix for Random Forest

Training:



Test:



Accuracies

Training Accuracy = 1.0

Test Accuracy = 1.0

Model 5: Boosting

About Boosting

Boosting is:

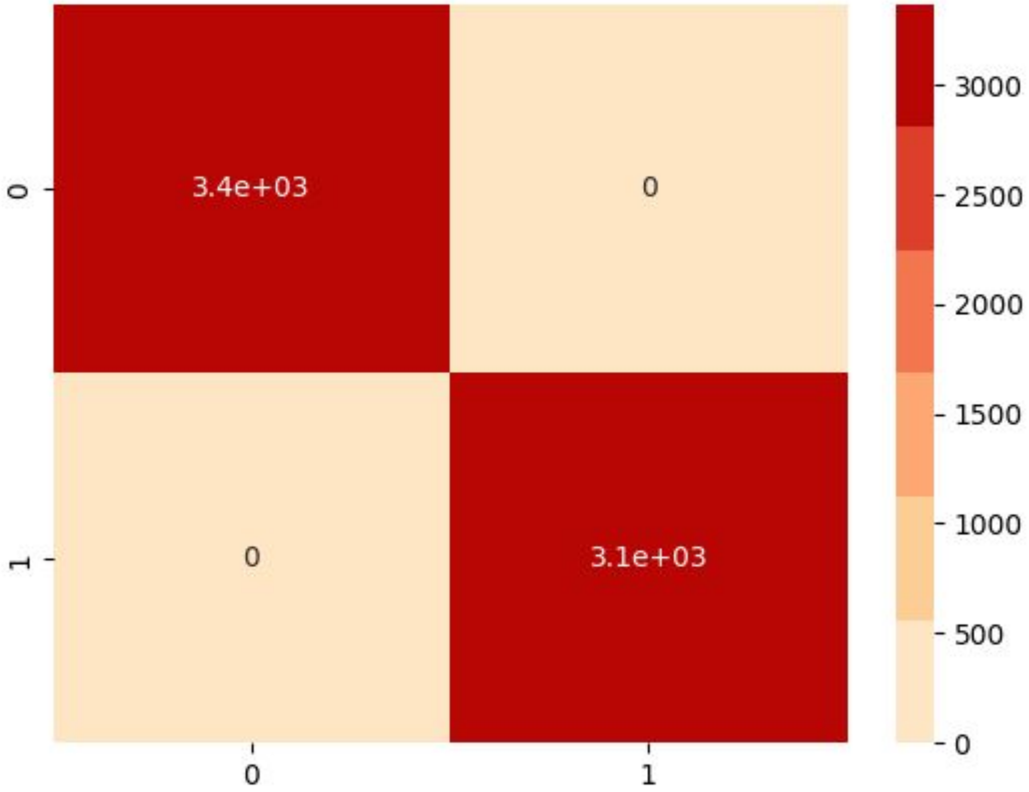
- **A NON-PARAMETRIC MODEL FOR BINARY CLASSIFICATION**
    - It relies on the aggregate performance of many “weak learner” models. Boosting assumes that each added new learner focuses on the errors made by previous learners.
- **BASED ON GRADIENT BOOSTING**

We Chose Boosting Because of its:

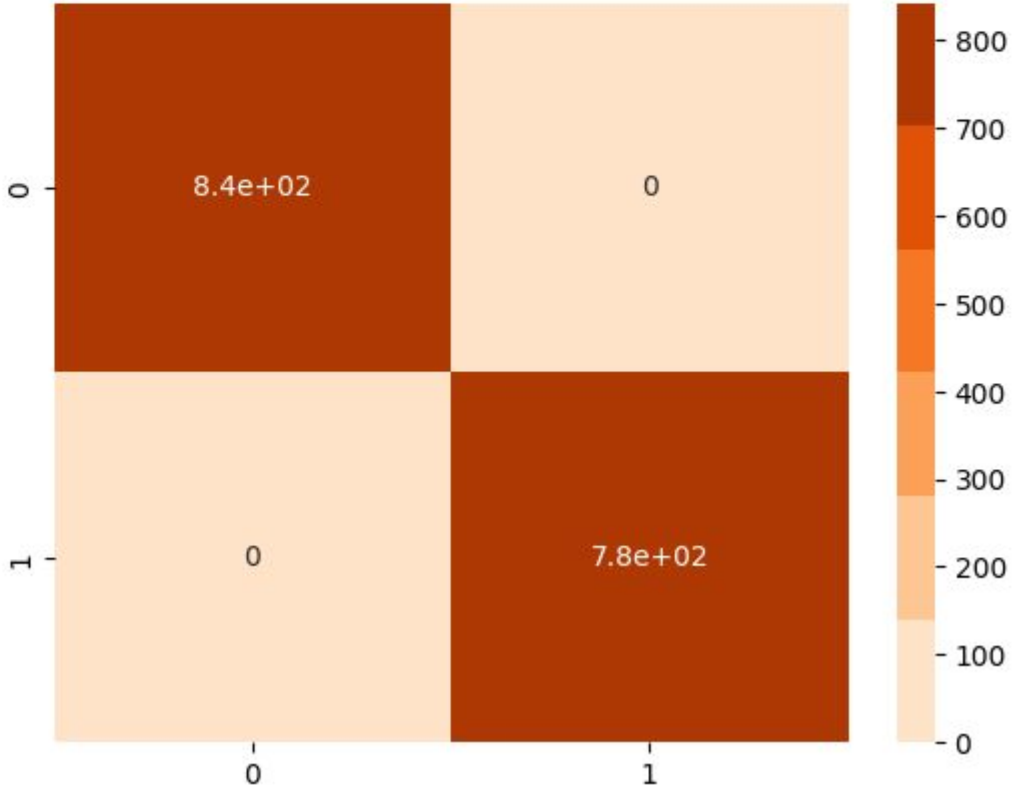
- **HIGH PREDICTIVE ACCURACY**
- **ABILITY TO HANDLE COMPLEX DATASETS**
- **HIGH EFFICIENCY AND SCALABILITY**
- **CAPABILITY OF MANAGING FEATURE INTERACTIONS**

Confusion Matrix for Boosting

Training:



Test:



Accuracies

Training Accuracy = 1.0
Test Accuracy = 1.0


Model 6: Support Vector Machine Classification (SVM)

About Support Vector Machines

SVM Classification is:

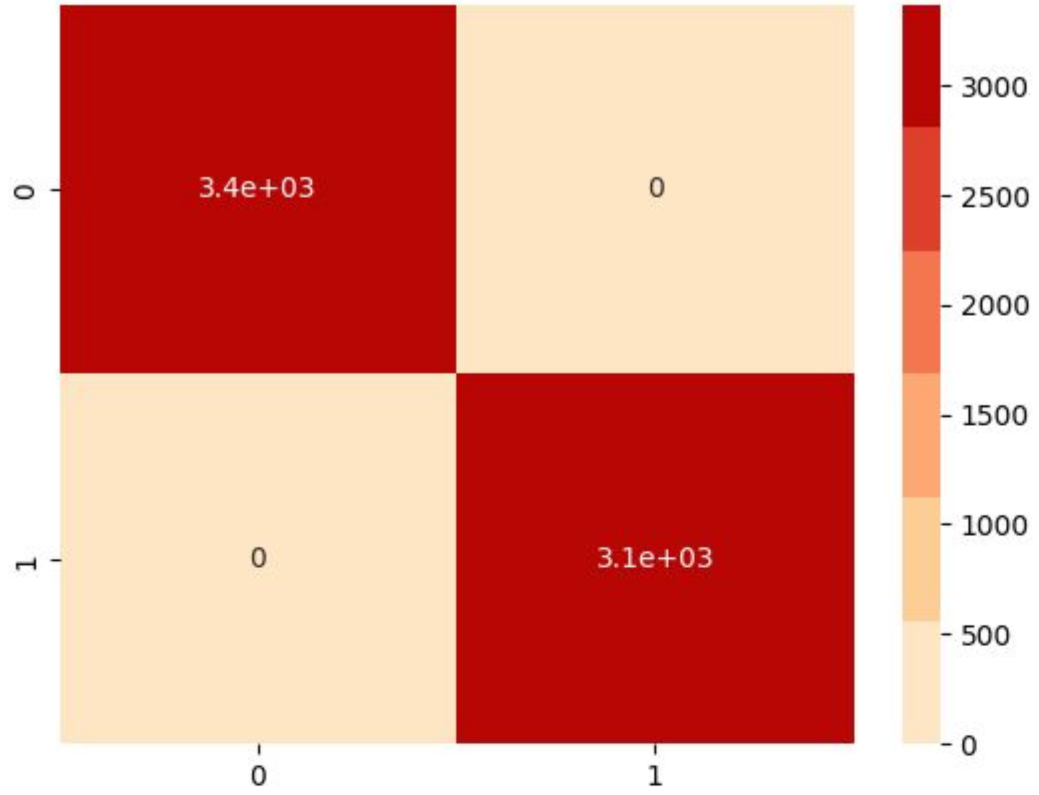
- **A PARAMETRIC MODEL**
  - Assumes the data is linearly separable or can be transformed into a higher-dimensional space where a linear separation exists
- **USED TO FIND THE OPTIMAL HYPERPLANE TO SEPARATE DIFFERENT CLASSES OF DATA**

We Chose SVM Because of its:

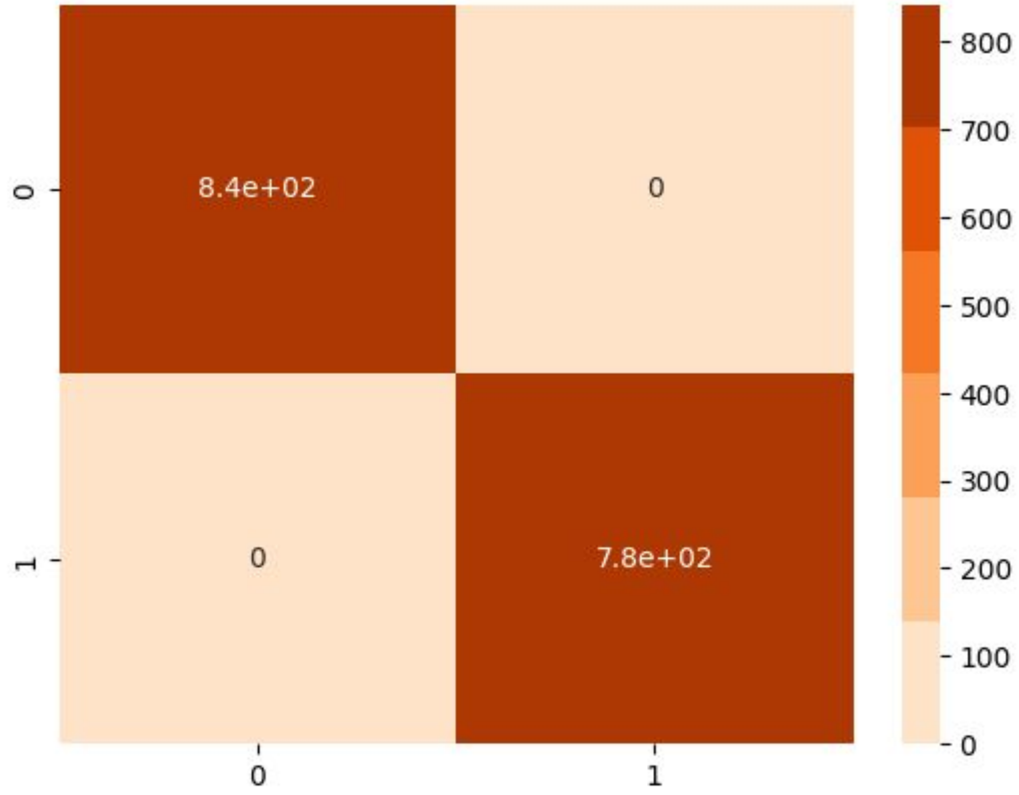
- **EFFECTIVENESS IN BINARY CLASSIFICATION**
- **ABILITY TO HANDLE NON-LINEAR RELATIONSHIPS IN THE DATA**

Confusion Matrix for SVM

Training:



Test:



Accuracies

Training Accuracy = 1.0
Test Accuracy = 1.0

Model 7: Naive-Bayes

About Naive-Bayes

Naive-Bayes is:

- **A PARAMETRIC MODEL**
  - Assumes that features are conditionally independent given the class label and following a Gaussian distribution
- **USED FOR BINARY CLASSIFICATION**
- **BASED UPON BAYES' THEOREM**

We Chose Naive-Bayes Because of its:

- **SIMPLICITY**
- **EFFICIENCY WITH HIGH-DIMENSIONAL DATA**
- **ABILITY TO PROVIDE NUMERICAL PROBABILISTIC PREDICTIONS**

Formulas for Naive-Bayes

Density of X given class j

$$f_j(\mathbf{x}) = f_{j1}(x_1) \times f_{j2}(x_2) \times \ldots \times f_{jp}(x_p)$$

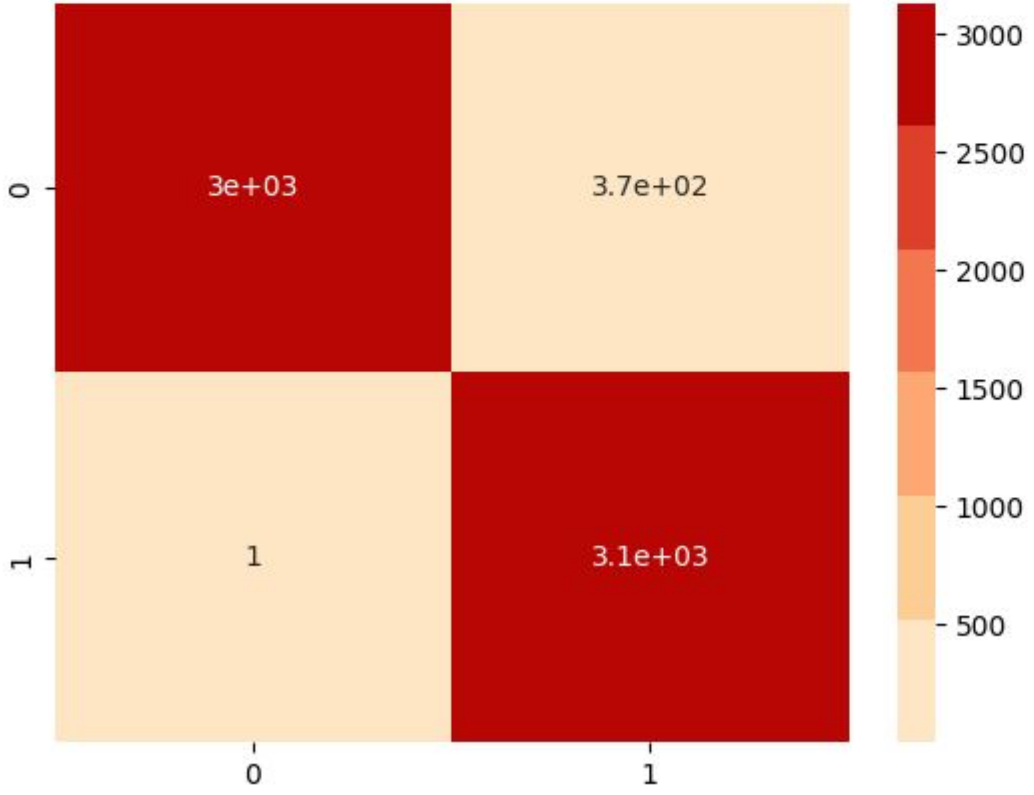
Bayes Theorem

$$p_j(\mathbf{x}) = \frac{\pi_j f_{j1}(x_1) \times f_{j2}(x_2) \times \ldots \times f_{jp}(x_p)}{\sum_j \pi_j f_{j1}(x_1) \times f_{j2}(x_2) \times \ldots \times f_{jp}(x_p)}$$

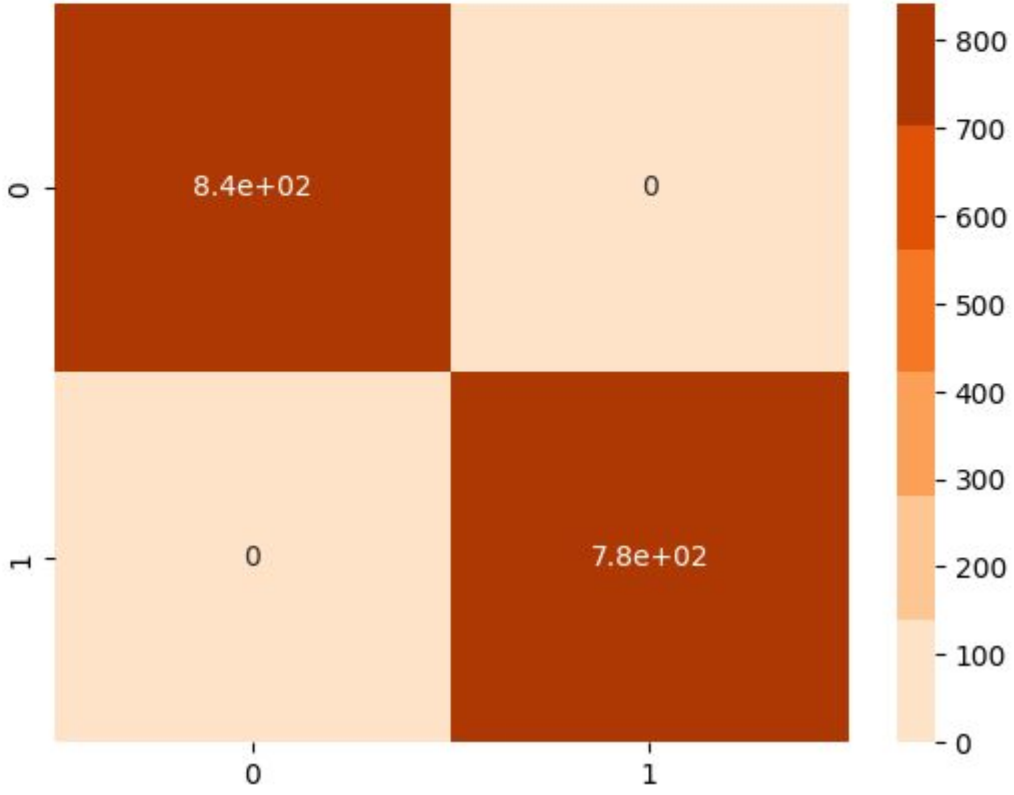
We choose the class with the largest probability computed from Bayes theorem

Confusion Matrix for Naive-Bayes

Training:



Test:

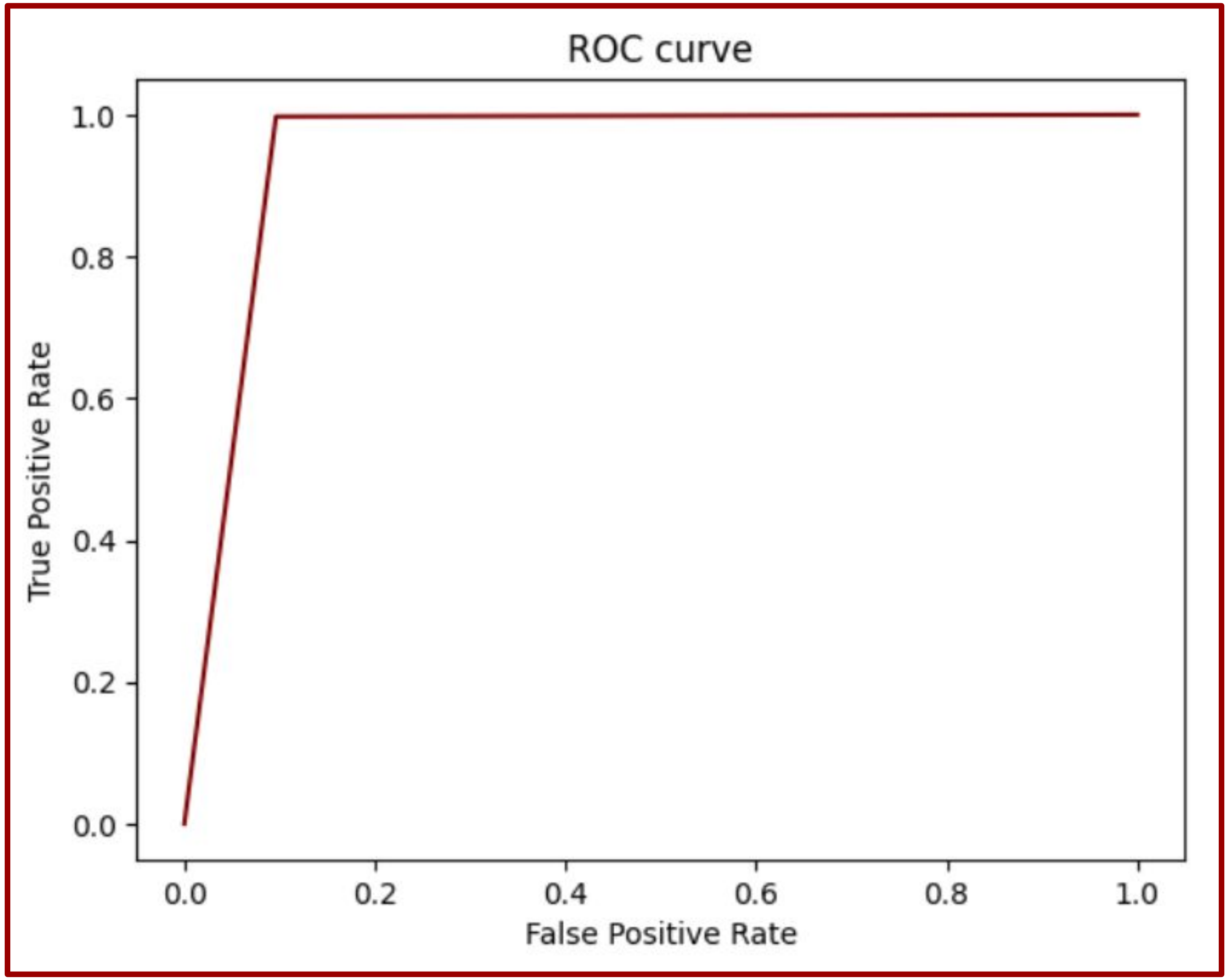


Accuracies

Training Accuracy = 0.9424526850284659

Test Accuracy = 0.9489230769230769

ROC Curve



AUC = 0.9506230983215176

Model 8: Linear Discriminant Analysis (LDA)

About Linear Discriminant Analysis

Linear Discriminant Analysis is:

- **A PARAMETRIC MODEL**
  - Assumes the normality and equality of covariance matrices of the features within each class. Focuses on maximizing the separation between classes.
- **USED FOR DIMENSIONALITY REDUCTION**

We Chose LDA Because of its:

- **EFFECTIVENESS IN FEATURE EXTRACTION**
- **INTERPRETABILITY**

Formulas for LDA

Discriminant Function

$$\delta_k(\boldsymbol{x}) = \tilde{\beta}_0^{(k)} + \tilde{\beta}_1^{(k)} x_1 + \ldots + \tilde{\beta}_p^{(k)} x_p$$

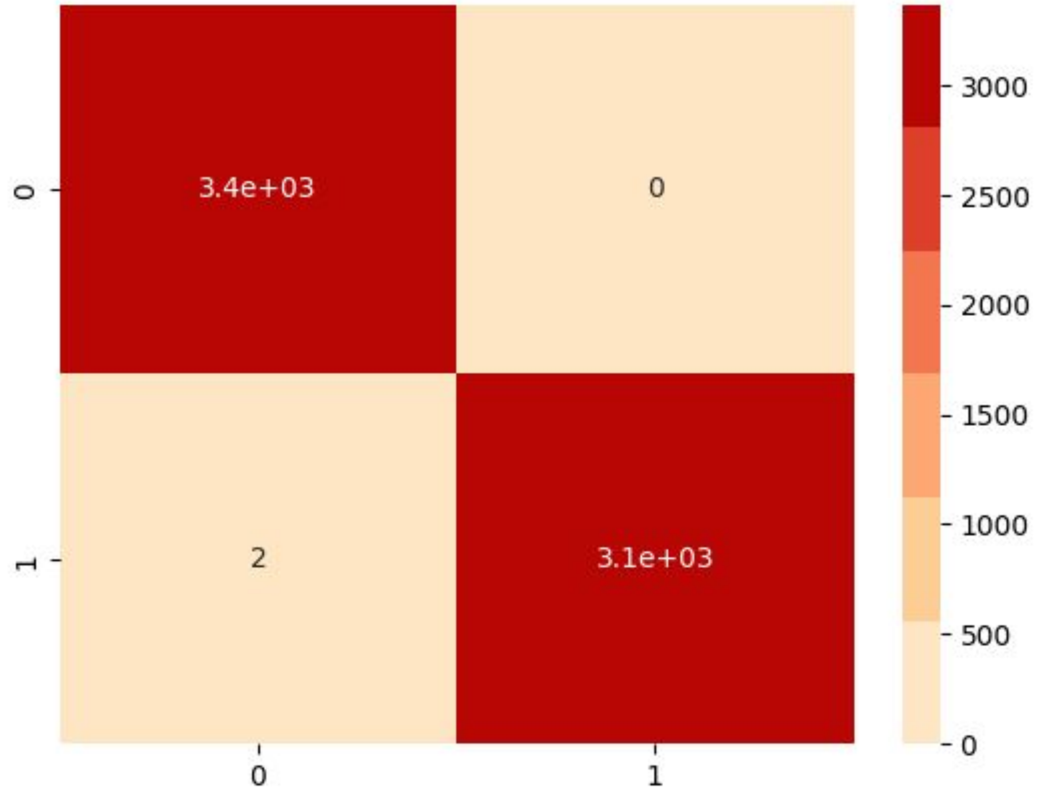
Posterior Probability of Classes

$$p_k(\boldsymbol{x}) = \frac{\pi_k f_k(\boldsymbol{x})}{\sum_{j=1}^{K} \pi_j f_j(\boldsymbol{x})} = \frac{e^{\delta_k(\boldsymbol{x})}}{\sum_{j=1}^{K} e^{\delta_j(\boldsymbol{x})}}$$

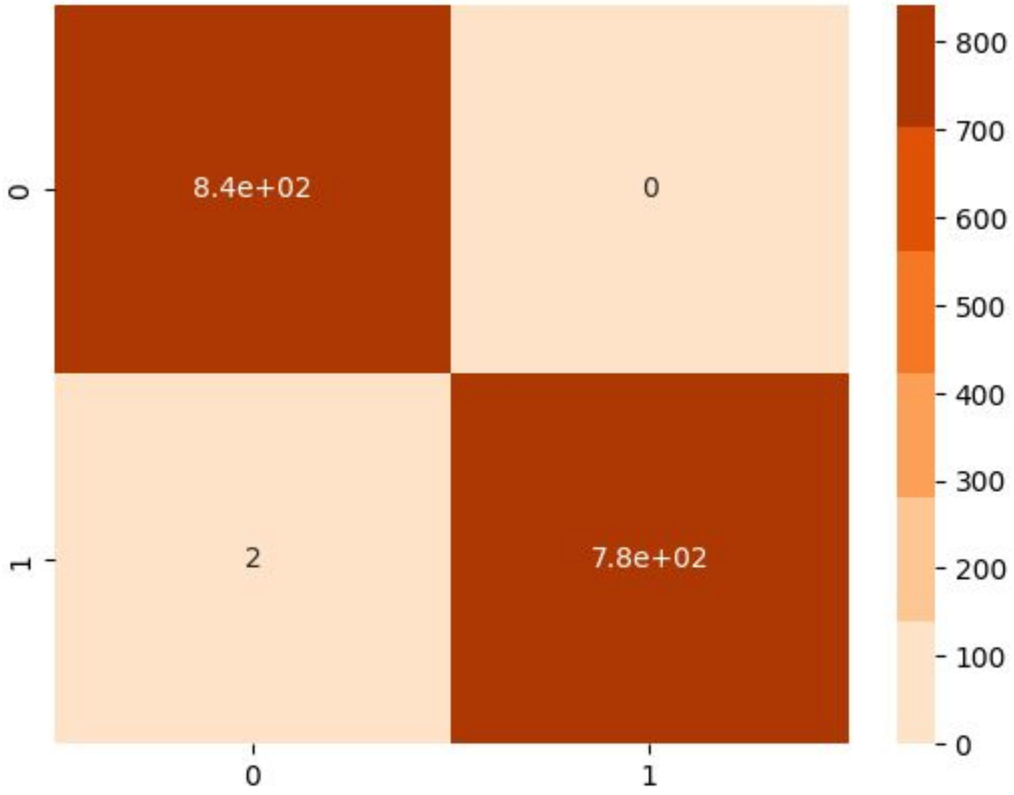
The predicted class has the largest discriminant function

Confusion Matrix for LDA

Training:



Test:



Accuracies

Training Accuracy = 0.9996922603477458

Test Accuracy = 0.9987692307692307

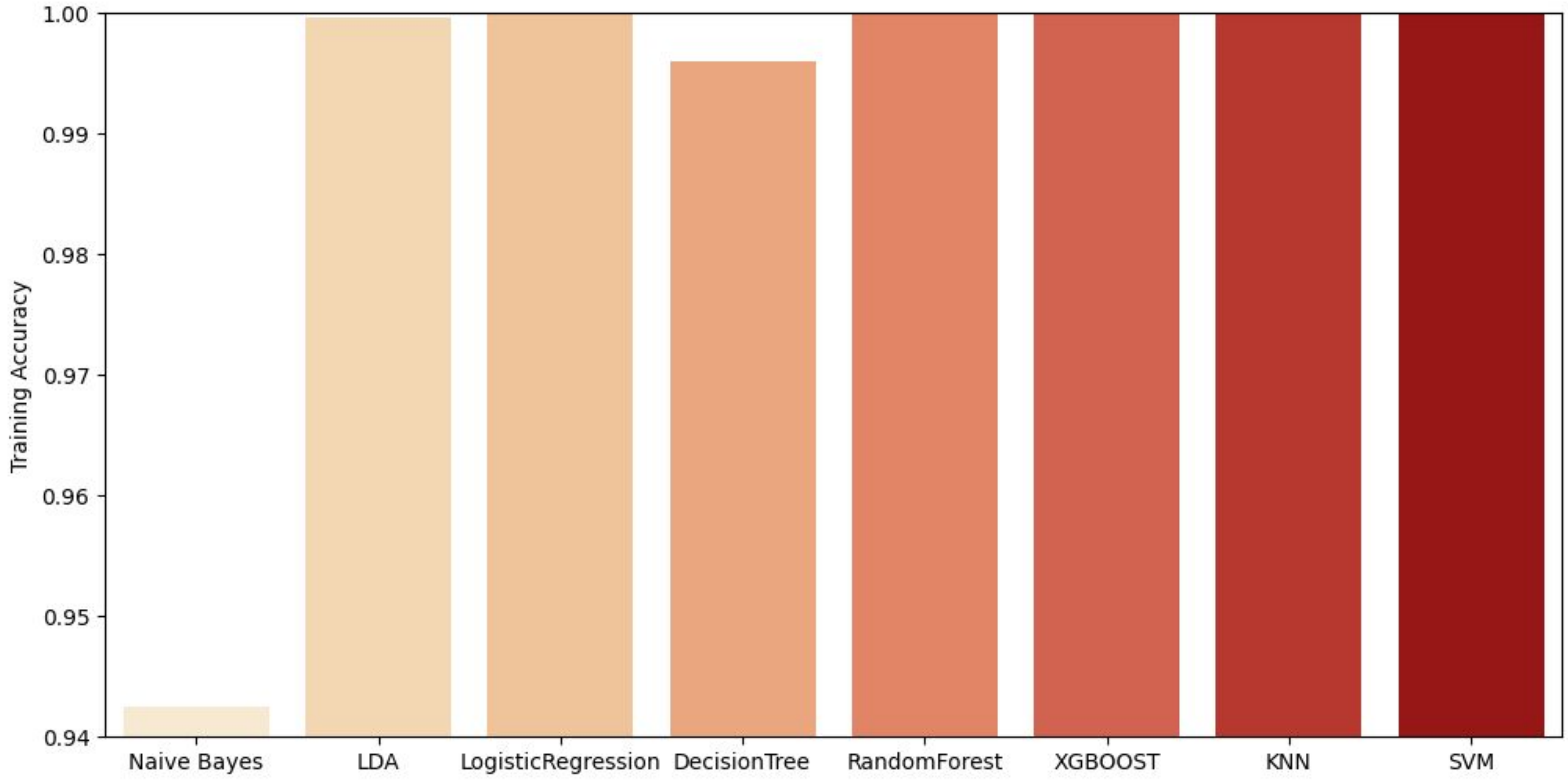
PART 06

Model Comparison

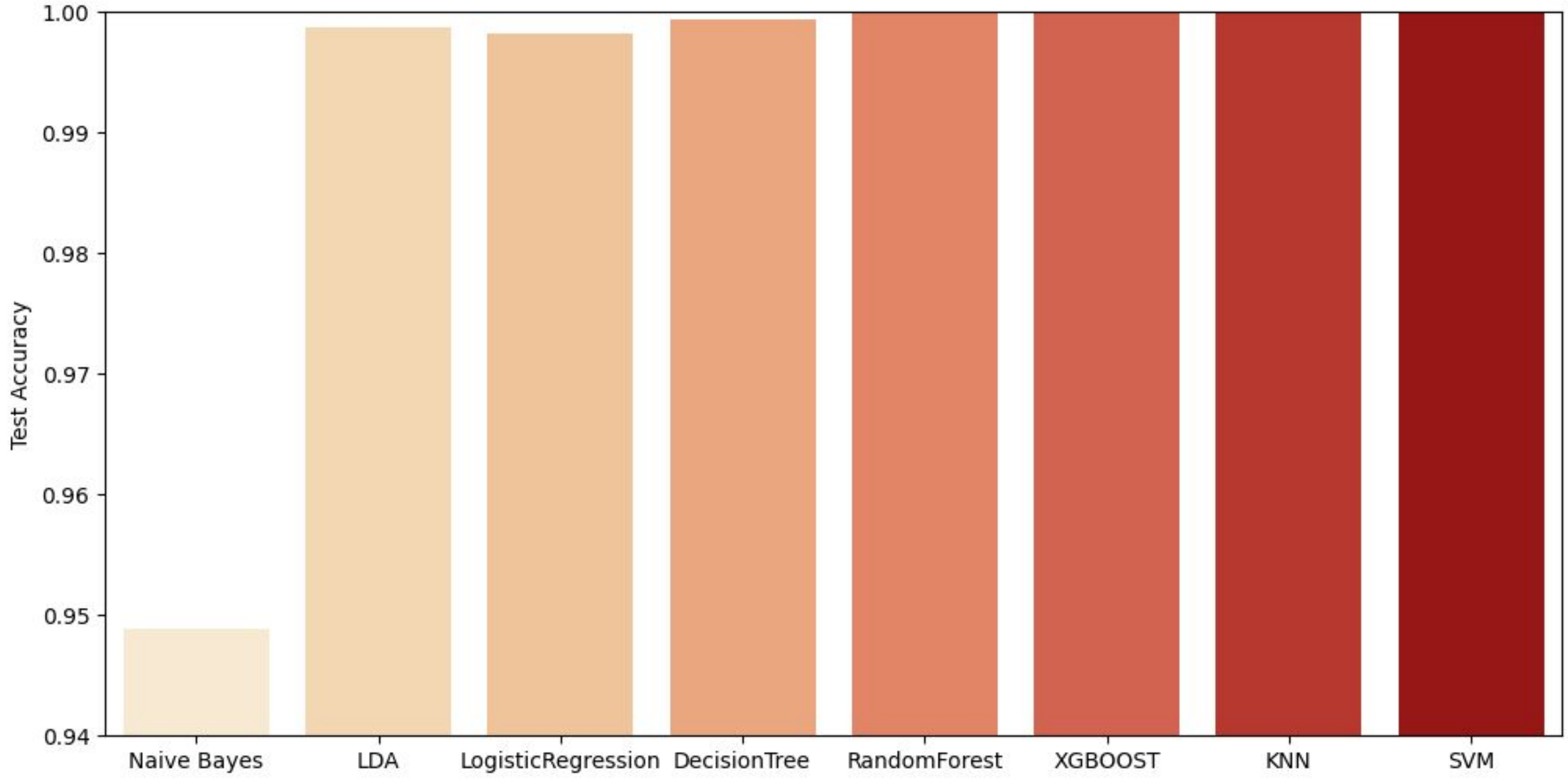
Comparing Accuracies For Our Models

Model	Training Accuracy	Test Accuracy	Training Mean	Test Mean
Naive Bayes	0.942453	0.948923	0.941068	0.911993
Decision Tree	0.995999	0.996307	0.995999	0.995690
LDA	0.999692	0.998769	0.999692	0.998152
LogisticRegression	0.999846	0.998154	0.999231	0.996921
RandomForest	1.000000	1.000000	1.000000	1.000000
XGBOOST	1.000000	1.000000	1.000000	1.000000
KNN	1.000000	1.000000	1.000000	1.000000
SVM	1.000000	1.000000	1.000000	1.000000

Histograms of Accuracies- Training



Histograms of Accuracies- Test



Why Are There Accuracies of 1?

Random Forest, Boosting, KNN, and SVM all have an accuracy score of 1

- The samples covered only 23 species of mushrooms, which were based off qualities in The Audubon Society Field Guide to North American Mushrooms
- Once our model accurately identifies what species a mushroom is, it can easily classify that species in the future as edible or poisonous
- This perfect classification rate could indicate some possible overfitting to our training data

Our Two Finalist Models:

01

We choose **Logistic Regression** for its:

- Interpretability
  - provides a clear linear decision boundary and its coefficients can be directly interpreted in terms of feature importance
- Easiness to Implement
- Efficiency for Computations
- Skill with Binary Classification Problems

02

We choose **Decision Trees** for their:

- Ease of Use in Real-life Scenarios
  - can be visually represented and therefore easily interpreted
- Interpretability
- Ability to Capture Complex Decision Boundaries

Most Influential Features for Logistic Regression

Inputs:

```
feature_importance = pd.Series(lr.coef_[0], index=x.columns)
print("Feature importance for Logistic Regression:")
z = feature_importance.sort_values(ascending=False)

for i in range(0,10): # Top 10
  feature_name = z.index[i-1]
  value = z[i]
  print(f"{feature_name} : {value}")
```

Outputs:

Feature importance for Logistic Regression:
odor_n : 4.336026937650609
spore-print-color_r : 3.431031183161751
odor_c : 3.254011744131949
gill-size : 3.0178596667878272
odor_p : 2.976425297438572
odor_f : 2.2959082907781347
stalk-root_b : 1.8343839906547967
stalk-surface-above-ring_k : 1.4976001673878978
stalk-surface-below-ring_y : 1.4830454517803968
population_c : 1.3269640301013133

Most Influential Features for Logistic Regression

Method Used

Analyzed the coefficients for each predictor variable, and noted the top 5 largest coefficients

Top 5 Most Influential Predictors

1. Odor
2. Spore Print Color
3. Gill Size
4. Stalk Root
5. Stalk Surface Above Ring

Most Influential Features for Decision Tree

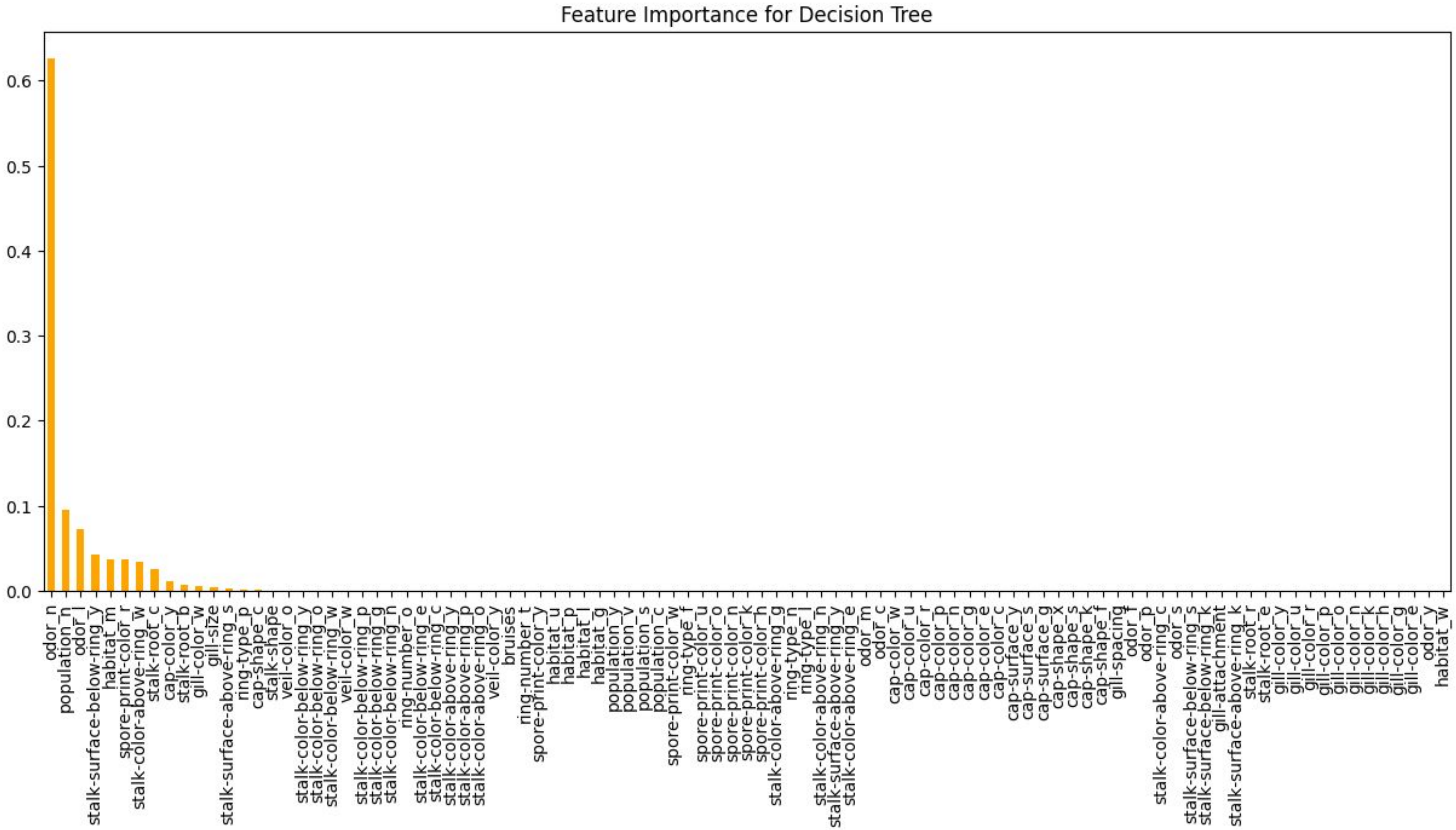
Inputs:

```
feature_importance =
pd.Series(dtree.feature_importances_,
index=x.columns)
print("Feature importance for Decision Tree:")
print(feature_importance.sort_values(ascending=F
alse))
```

Outputs:

```
Feature importance for Decision Tree:
odor_n                        0.225838
ring-type_l                   0.212852
spore-print-color_w           0.169990
odor_f                        0.107429
spore-print-color_k           0.092850
                 ...
gill-color_o                  0.000000
gill-color_n                  0.000000
gill-color_k                  0.000000
gill-color_h                  0.000000
habitat_w                     0.000000
Length: 95, dtype: float64
```

Most Influential Features for Decision Tree



| Most Influential Features for Decision Tree

Method Used

Implemented the feature importance function for decision trees and selected the top 5 ranked variables

Top 5 Most Influential Predictors

1. Odor
2. Population
3. Stalk Surface Below Ring
4. Habitat
5. Spore Print Color

Common Influential Features

Common Influential Features Shared by Logistic Regression & Decision Trees are:

ODOR

SPORE PRINT COLOR

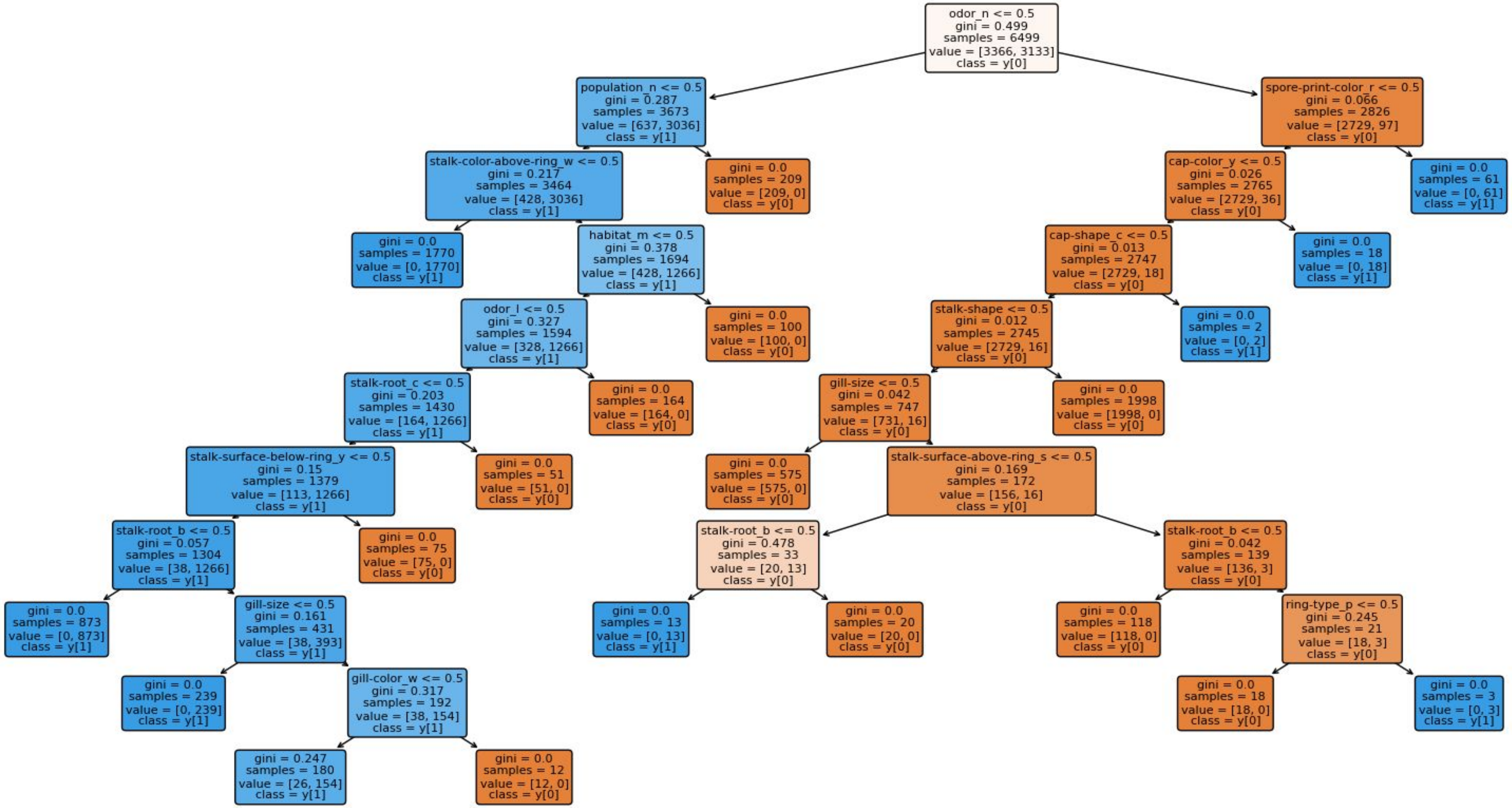
Our Final Chosen Model is...

DECISION TREE

Why?

- It requires no computational background to use as it is a visual representation of classification with clear decision boundaries
- It is a realistic and simple way to classify mushrooms for the average user
- Our decision tree model showed high performance accuracy

Our Final Chosen Model is...



PART 07

Real-Life Sample

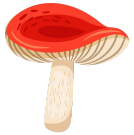
Our Sample

We obtained a mushroom of the species **Agaricus Bisporus**, commonly known as the **Portobello mushroom**, and inputted its qualities into each of our models.

- Since we know this mushroom is edible, it provides a perfect opportunity to test our model's classification accuracies in real life

| Inputs- Portobello Mushroom

```
data_dict = {
  'bruises': 1,
  'gill-attachment': 1,
  'gill-spacing': 0,
  'gill-size': 1,
  'stalk-shape': 1,
  'cap-shape_c': 0,
  'cap-shape_f': 0,
  'cap-shape_k': 0,
  'cap-shape_s': 0,
  'cap-shape_x': 1,
  'cap-surface_g': 0,
  'cap-surface_s': 1,
  'cap-surface_y': 0,
  'cap-color_c': 0,
  'cap-color_e': 0,
  'cap-color_g': 0,
  'cap-color_n': 1,
  'cap-color_p': 0,
  'cap-color_r': 0,
  'cap-color_u': 0,
  'cap-color_w': 0,
  'cap-color_y': 0,
  'odor_c': 0,
  'odor_f': 0,
  'odor_l': 0,
  'odor_m': 0,
  'odor_n': 1,
  'odor_p': 0,
  'odor_s': 0,
  'odor_y': 0,
  'gill-color_e': 0,
  'gill-color_g': 0,
  'gill-color_h': 0,
  'gill-color_k': 0,
  'gill-color_n': 1,
  'gill-color_o': 0,
  'gill-color_p': 0,
  'gill-color_r': 0,
  'gill-color_u': 0,
  'gill-color_w': 0,
  'gill-color_y': 0,
  'stalk-root_b': 0,
  'stalk-root_c': 0,
  'stalk-root_e': 0,
  'stalk-root_r': 0,
  'stalk-surface-above-ring_k': 0,
  'stalk-surface-above-ring_s': 1,
  'stalk-surface-above-ring_y': 0,
  'stalk-surface-below-ring_k': 0,
  'stalk-surface-below-ring_s': 1,
  'stalk-surface-below-ring_y': 0,
  'stalk-color-above-ring_c': 0,
  'stalk-color-above-ring_e': 0,
  'stalk-color-above-ring_g': 0,
  'stalk-color-above-ring_n': 1,
  'stalk-color-above-ring_o': 0,
  'stalk-color-above-ring_p': 0,
  'stalk-color-above-ring_w': 0,
  'stalk-color-above-ring_y': 0,
  'stalk-color-below-ring_c': 0,
  'stalk-color-below-ring_e': 0,
  'stalk-color-below-ring_g': 0,
  'stalk-color-below-ring_n': 1,
  'stalk-color-below-ring_o': 0,
  'stalk-color-below-ring_p': 0,
  'stalk-color-below-ring_w': 0,
  'stalk-color-below-ring_y': 0,
  'veil-color_o': 0,
  'veil-color_w': 1,
  'veil-color_y': 0,
  'ring-number_o': 1,
  'ring-number_t': 0,
  'ring-type_f': 0,
  'ring-type_l': 0,
  'ring-type_n': 1,
  'ring-type_p': 0,
  'spore-print-color_h': 0,
  'spore-print-color_k': 0,
  'spore-print-color_n': 1,
  'spore-print-color_o': 0,
  'spore-print-color_r': 0,
  'spore-print-color_u': 0,
  'spore-print-color_w': 0,
  'spore-print-color_y': 0,
  'population_c': 0,
  'population_n': 1,
  'population_s': 0,
  'population_v': 0,
  'population_y': 0,
  'habitat_g': 0,
  'habitat_l': 0,
  'habitat_m': 0,
  'habitat_p': 0,
  'habitat_u': 0,
  'habitat_w': 0
```



Results

Model	Prediction
Logistic Regression	0
KNN	0
Decision Tree	0
Random Forest	0
Boosting	0
SVM	0
Naive-Bayes	1
LDA	0

- After inputting the qualities of Portobello mushrooms into our models, we obtained a classification from each model indicating whether they are edible or poisonous
- Every model returned a value of 0 (indicating edibility), except for Naive-Bayes
    - Naive-Bayes has the lowest accuracy of all of our models, so this makes sense
- So, our models are applicable to the real world!

PART 08

Conclusion

Key Takeaways

01 There is no shortcut to determining the edibility of mushrooms, either extensive knowledge or a machine learning model is necessary

02 Decision Trees and Logistic Regression models ultimately fit our needs for a model the best, as they prioritize accuracy and ease of use

03 When simply looking at a mushroom of unknown edibility, we recommend examining the odor and spore print color, as they are the two most influential predictors on edibility

04 With high accuracy scores, our models indicate excellent model performance, however there may be some overfitting due to our samples coming only from 23 mushroom species

Thank you!

Any Questions?

Works Cited

Banach, Gail. “Poison Prevention - SUNY Upstate Medical University.” *Poison Prevention*, 2021, www.upstate.edu/poison/pdf/pp_newsletter/2021/june-2021-newsletter.pdf.

Barnhart, Max. “New Hope for an Antidote to Death Cap Mushrooms and Other Poison Fungi.” *NPR*, NPR, 17 May 2023, www.npr.org/sections/goatsandsoda/2023/05/17/1175494500/new-hope-for-an-antidote-to-death-cap-mushrooms-and-other-poison-fungi#:~:text=Eating%20only%20half%20a%20cap,100%20deaths%20a%20year%20globally.

Dturck. “Oddities and Fun Facts about Agaricus Bisporus.” *A Fungus Among Us*, 19 Mar. 2018, honorsfungi.wordpress.com/2018/03/19/oddities-and-fun-facts-about-agaricus-bisporus/#:~:text=Agaricus%20bisporus%20accounts%20for%20around,the%20white%20and%20button%20varieties.

Eren, Sevki Hakan, et al. “Mushroom Poisoning: Retrospective Analysis of 294 Cases.” *Clinics (Sao Paulo, Brazil)*, U.S. National Library of Medicine, May 2010, www.ncbi.nlm.nih.gov/pmc/articles/PMC2882543/#:~:text=Various%20types%20of%20wild%20mushrooms,3%25%20of%20these%20are%20poisonous.

Works Cited (Continued)

"Global Distribution of the Main Production Area 's of Button Mushrooms ..." *Research Gate*, www.researchgate.net/figure/Global-distribution-of-the-main-production-area-s-of-button-mushrooms-Agaricus_fig1_283069137. Accessed 1 Dec. 2023.

Jenny. "Unmasking the 7 Deadliest Poisonous Mushroom Species in the US." *Mushroom Appreciation*, 30 June 2023, www.mushroom-appreciation.com/7-deadliest-poisonous-mushroom-species.html#:~:text=There%20are%20thousands%20of%20mushroom, severe%20illness%20or%20even%20death.

"Lepiota." *Wikipedia*, Wikimedia Foundation, 25 Oct. 2023, en.wikipedia.org/wiki/Lepiota.

Lincoff, Gary, and Carol Nehring. "National Audubon Society Field Guide to North American Mushrooms." *Amazon*, Knopf, 1997, www.amazon.com/National-Audubon-Society-American-Mushrooms/dp/0394519922.

Lincoff, Gary H. The Audubon Society Field Guide to North American Mushrooms. Knopf, 1981.

"Mushroom Classification." Kaggle, UCI Machine Learning, 1 Dec. 2016, www.kaggle.com/datasets/uciml/mushroom-classification/data.

Petruzzello, Melissa. "Portobello Mushroom." Encyclopædia Britannica, Encyclopædia Britannica, inc., 31 May 2023, www.britannica.com/topic/portobello-mushroom.

Ria, Nushrat, et al. ICCCNT, 2021, pp. 1–5, State of Art Research in Edible and Poisonous Mushroom Recognition.