Mushroom Classification

By Morgan Keller and Tyson Hyland

Purpose

To classify mushrooms as edible or poisonous

Misclassifying edible as poisonous is not good, but relatively safe, while misclassifying poisonous as edible is potentially fatal

True Class

	Positive	Negative
Positive	TP	FP
Negative	FN	TN

Predicted Class

Data – 52% Poisonous, 48% Edible – None Missing

class	cap-shape	cap-surface	cap-color	bruises	odor	gill-attach	gill-spacing	gill-size	gill-color	stalk-shape	stalk-root	stalk-surface	stalk-surface	stalk-color	stalk-color	veil-type	veil-color	ring-number	ring-type	spore-print	population	habitat
p	x	s	n	t	p	f	c	n	k	e	e	s	s	w	w	p	w	o	p	k	s	u
e	x	s	y	t	a	f	c	b	k	e	c	s	s	w	w	p	w	o	p	n	n	g
e	b	s	w	t	l	f	c	b	n	e	c	s	s	w	w	p	w	o	p	n	n	m
p	x	y	w	t	p	f	c	n	n	e	e	s	s	w	w	p	w	o	p	k	s	u
e	x	s	g	f	n	f	w	b	k	t	e	s	s	w	w	p	w	o	e	n	a	g
e	x	y	y	t	a	f	c	b	n	e	c	s	s	w	w	p	w	o	p	k	n	g
e	b	s	w	t	a	f	c	b	g	e	c	s	s	w	w	p	w	o	p	k	n	m
e	b	y	w	t	l	f	c	b	n	e	c	s	s	w	w	p	w	o	p	n	s	m
p	x	s	w	t	p	f	c	n	p	e	e	s	s	w	w	p	w	o	p	k	v	g
e	b	s	y	t	a	f	c	b	g	e	c	s	s	w	w	p	w	o	p	k	s	m
e	x	y	y	t	l	f	c	b	g	e	c	s	s	w	w	p	w	o	p	n	n	g
e	x	y	y	t	a	f	c	b	n	e	c	s	s	w	w	p	w	o	p	k	s	m
e	b	s	y	t	a	f	c	b	w	e	c	s	s	w	w	p	w	o	p	n	s	g
p	x	y	w	t	p	f	c	n	k	e	e	s	s	w	w	p	w	o	p	n	v	u
e	x	f	n	f	n	f	w	b	n	t	e	s	f	w	w	p	w	o	e	k	a	g
e	s	f	g	f	n	f	c	n	k	e	e	s	s	w	w	p	w	o	p	n	y	u
f	f	f	w	f	n	f	w	b	k	t	e	s	s	w	w	p	w	o	e	n	a	g
p	x	s	n	t	p	f	c	n	k	e	e	s	s	w	w	p	w	o	p	k	s	g
e	x	y	w	t	p	f	c	n	n	e	e	s	s	w	w	p	w	o	p	n	s	u
p	x	s	n	t	p	f	c	n	k	e	e	s	s	w	w	p	w	o	p	n	s	u
e	b	s	y	t	a	f	c	b	k	e	c	s	s	w	w	p	w	o	p	n	s	m
p	x	y	n	t	p	f	c	n	n	e	e	s	s	w	w	p	w	o	p	n	v	g

Preprocessing Data - Label Encoding

```
Class label mapping: {'e': 0, 'p': 1}

cap-shape label mapping: {'b': 0, 'c': 1, 'f': 2, 'k': 3, 's': 4, 'x': 5}

cap-surface label mapping: {'f': 0, 'g': 1, 's': 2, 'y': 3}

Cap-color label mapping: {'b': 0, 'c': 1, 'e': 2, 'g': 3, 'n': 4, 'p': 5, 'r': 6, 'u': 7, 'w': 8, 'y': 9}

bruises label mapping: {'f': 0, 't': 1}

Odor label mapping: {'a': 0, 'c': 1, 'f': 2, 'l': 3, 'm': 4, 'n': 5, 'p': 6, 's': 7, 'y': 8}

gill-attachment label mapping: {'a': 0, 'f': 1}

gill-spacing label mapping: {'c': 0, 'w': 1}

gill-size label mapping: {'b': 0, 'n': 1}

gill-color label mapping: {'b': 0, 'e': 1, 'g': 2, 'h': 3, 'k': 4, 'n': 5, 'o': 6, 'p': 7, 'r': 8, 'u': 9, 'w': 10, 'y': 11}

stalk-shape label mapping: {'e': 0, 't': 1}

stalk-root label mapping: {'?': 0, 'b': 1, 'c': 2, 'e': 3, 'r': 4}

stalk-surface-above-ring label mapping: {'f': 0, 'k': 1, 's': 2, 'y': 3}

stalk-surface-below-ring label mapping: {'f': 0, 'k': 1, 's': 2, 'y': 3}

stalk-color-above-ring label mapping: {'b': 0, 'c': 1, 'e': 2, 'g': 3, 'n': 4, 'o': 5, 'p': 6, 'w': 7, 'y': 8}

stalk-color-below-ring label mapping: {'b': 0, 'c': 1, 'e': 2, 'g': 3, 'n': 4, 'o': 5, 'p': 6, 'w': 7, 'y': 8}

veil-type label mapping: {'p': 0}

veil-color label mapping: {'n': 0, 'o': 1, 'w': 2, 'y': 3}

ring-number label mapping: {'n': 0, 'o': 1, 't': 2}

ring-type label mapping: {'e': 0, 'f': 1, 'l': 2, 'n': 3, 'p': 4}

spore-print-color label mapping: {'b': 0, 'h': 1, 'k': 2, 'n': 3, 'o': 4, 'r': 5, 'u': 6, 'w': 7, 'y': 8}

population label mapping: {'a': 0, 'c': 1, 'n': 2, 's': 3, 'v': 4, 'y': 5}

habitat label mapping: {'d': 0, 'g': 1, 'l': 2, 'm': 3, 'p': 4, 'u': 5, 'w': 6}
```

Methods and Stuff We Did

- Chi2 Test
- Histogram Plots
- Decision Trees
- Naive Bayes Update Odds

	Variable	Chi2	P-value
8	gill-color	5957.764469	0.000000e+00
18	ring-type	1950.610146	0.000000e+00
7	gill-size	1636.606833	0.000000e+00
3	bruises	1194.277352	1.069037e-261
10	stalk-root	1186.029221	6.630617e-260
6	gill-spacing	826.795274	8.061293e-182
21	habitat	751.309489	2.082916e-165
19	spore-print-color	379.132729	1.922119e-84
20	population	311.766736	9.002366e-70
11	stalk-surface-above-ring	222.982400	2.022392e-50
1	cap-surface	214.068544	1.779362e-48
12	stalk-surface-below-ring	206.648180	7.398869e-47
13	stalk-color-above-ring	119.792216	7.024638e-28
14	stalk-color-below-ring	109.789410	1.089740e-25
4	odor	75.910163	2.968682e-18
9	stalk-shape	36.594105	1.454709e-09
17	ring-number	25.646335	4.100735e-07
0	cap-shape	17.508364	2.860465e-05
2	cap-color	11.511382	6.917135e-04
16	veil-color	5.126826	2.355878e-02
5	gill-attachment	3.505447	6.116734e-02
15	veil-type	NaN	NaN

Decision Trees - Hypothesis

5 features:

Odor,

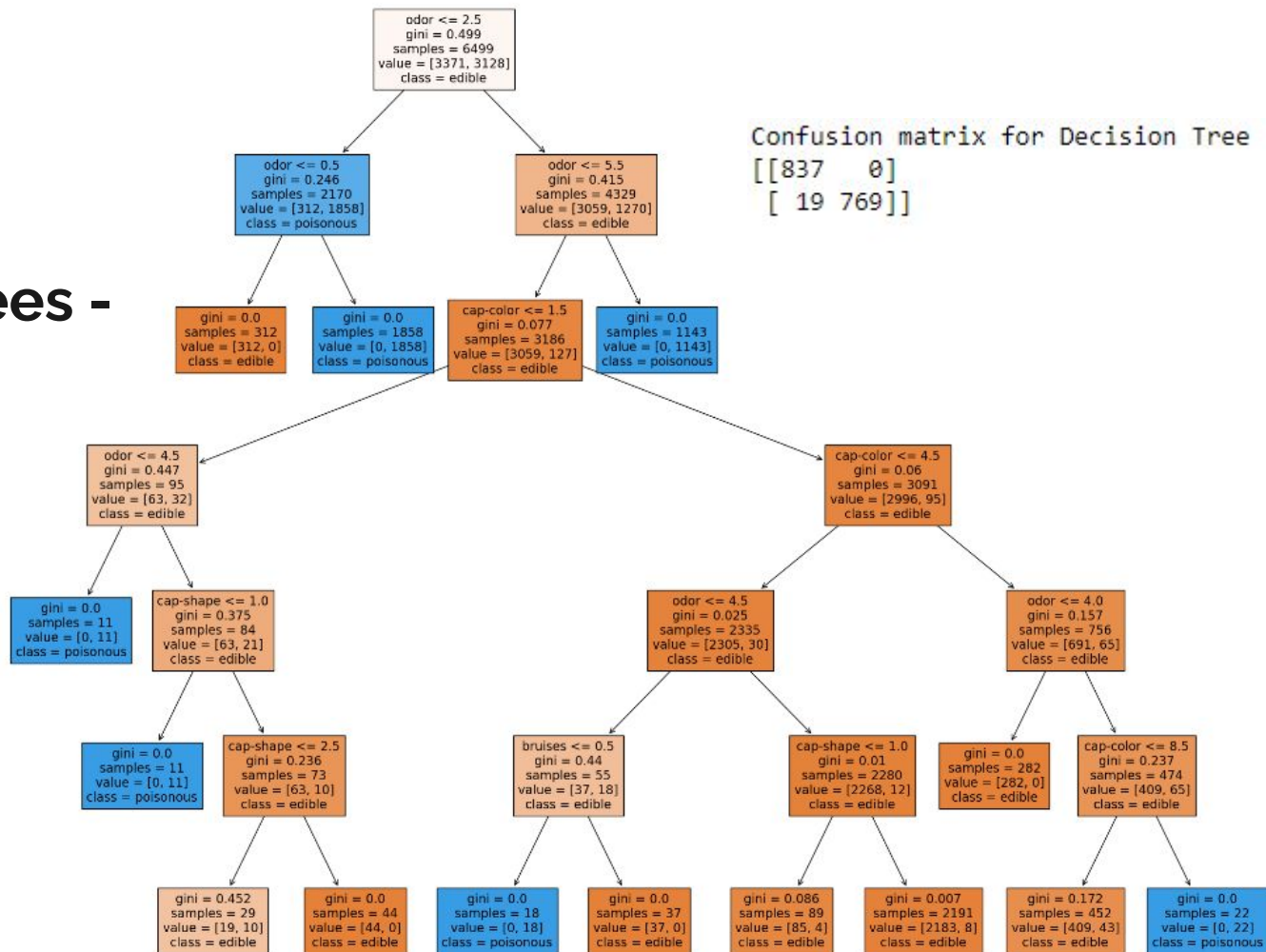Cap Color,

Cap Shape,

Cap Surface,

Bruises

MaxDepth = 6

Confusion matrix for Decision Tree
```
[[837    0]
 [ 19  769]]
```
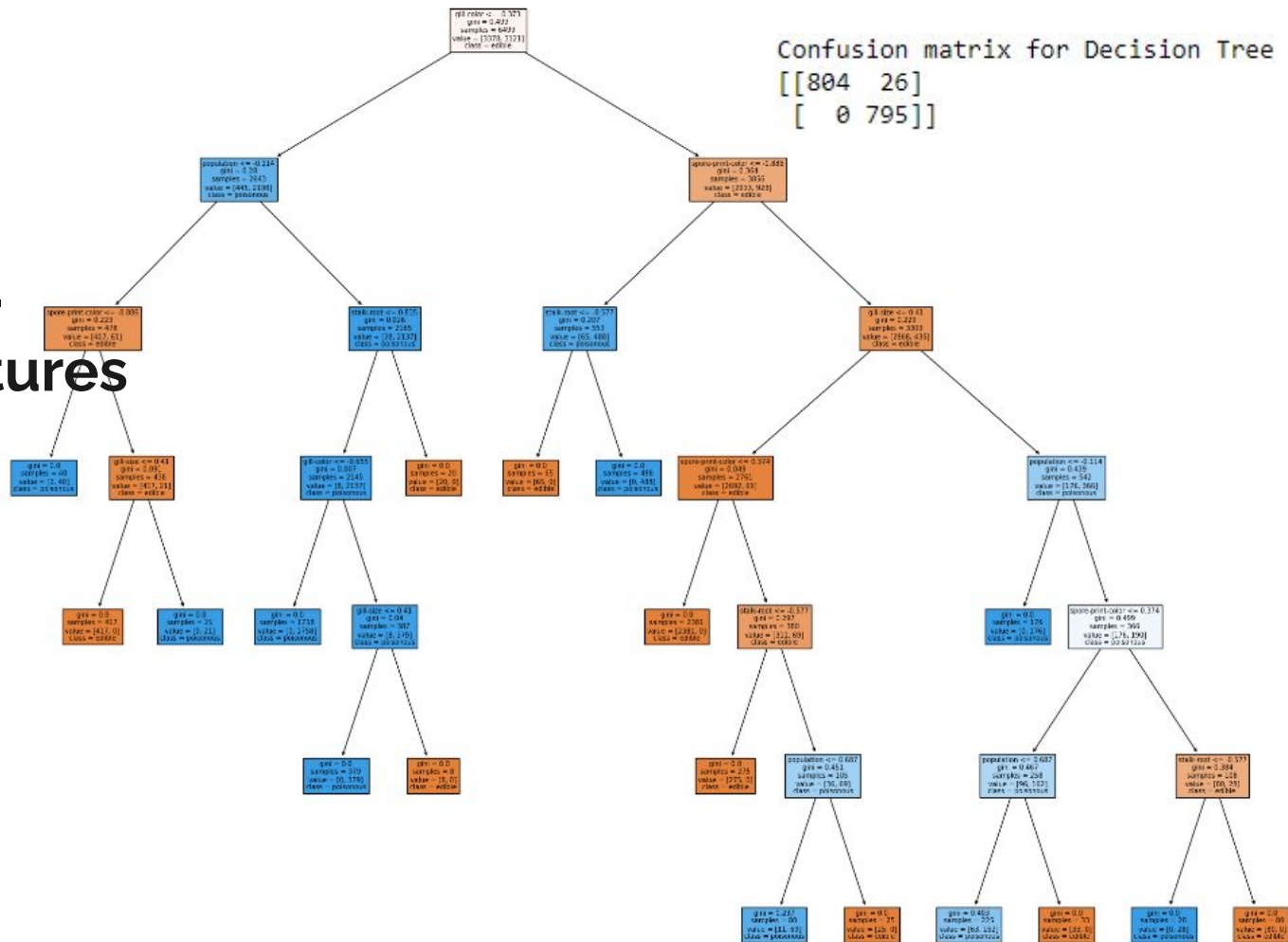
Decision Trees- RFE 5 Best Features

Gill Size,

Gill Color,

Stalk root,

Spore print color,

Population

MaxDepth = 6

Confusion matrix for Decision Tree
[[804 26]
 [0 795]]

Naive Bayes