$1 - \frac{1}{36} - \frac{25}{36} = 1 - \frac{26}{36} = \frac{10}{36} = \frac{5}{18}$

Brian Holliday
Professor C
Intro to Data Mining
30 March 2020

Project #5 Part 1

GINI Index

(1)

B	+	−	P(t_i)	GINI(t_i)
t_1 = T	3	1	4/10	$1 - (3/4)^2 - (1/4)^2 = 3/8$
t_2 = F	1	5	6/10	$1 - (1/6)^2 - (5/6)^2 = 5/18$

(2) $GINI\ split = \sum_{i=1}^{n} P(t_i) \cdot GINI(t_i) = (4/10)(3/8) + (6/10)(5/18)$

$= 0.31667$

A	+	−	P(t_i)	GINI(t_i)
t_1 = T	4	3	7/10	$1 - (4/7)^2 - (3/4)^2 = 25/49$
t_2 = F	0	3	3/10	$1 - (0/3)^2 - (3/3)^2 = 0$

$GINI\ split = \sum P(t_i) \cdot GINI(t_i) = (7/10)(25/49) + (3/10)(0)$

$= \frac{175}{490} = 0.3514$

We would choose B because the GINI Index is smaller therefore closer to having pure nodes

Misclassification Error

(2)

B	+	−	P(t_i)	Error(t_i)
t_1 = T	3	1	4/10	$1 - 3/4 = 1/4$
t_2 = F	1	5	6/10	$1 - 5/6 = 1/6$

Error Split = $(4/10)(1/4) + (6/10)(1/6) =$

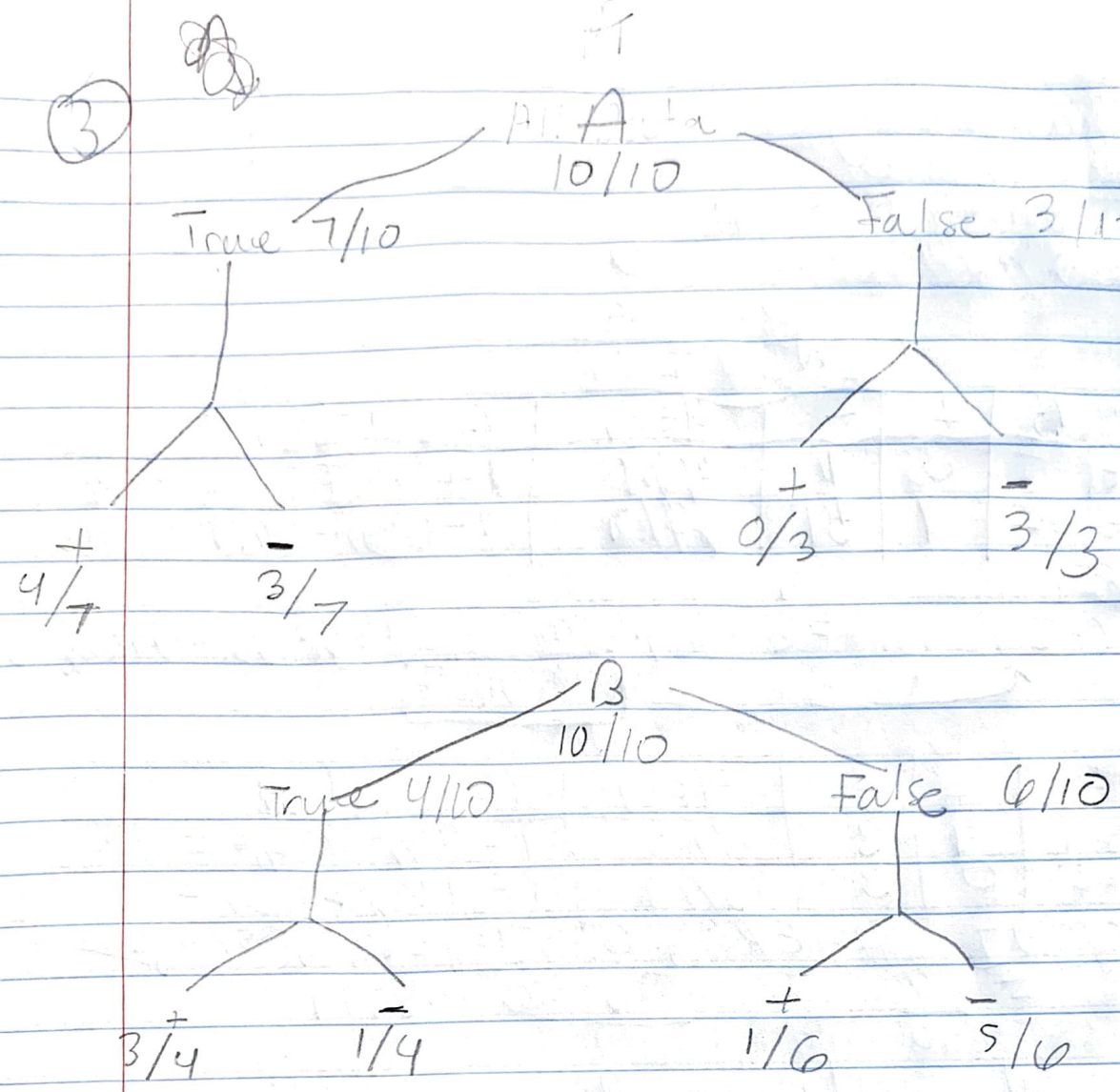
$= 0.36667$

A	+	−	P(t_i)	Error(t_i)
t_1 = T	4	3	7/10	$1 - (3/7) = 4/7$
t_2 = F	0	3	3/10	$1 - 3/3 = 0$

Error = $(7/10)(4/7) + (3/10)(0) = 0.40$

Based on the classification error we would choose B

(3)


A
10/10
True 7/10
False 3/10
+
4/7
-
3/7
+
0/3
-
3/3
B
10/10
True 4/10
False 6/10
+
3/4
-
1/4
+
1/6
-
5/6