TOY DATA

ID	DEP DEL 15	PREV DEL DEL	AIR TIME GAP
1	1	1	15
2	1	0	2
3	1	0	35
4	0	0	40
5	0	1	68
6	0	0	30
7	0	0	90
8	0	0	28

THE SOURCE CODE FOR THE GBTClassifier() FUNCTION INDICATES THAT VARIANCE IS THE IMPURITY MEASURE

IMPURITY (BASE) = var $[1,1,1,0,0,0,0,0]$ = 0.268

PREV DEP DEL

0	0	0		0	0	0		1	1
0	0	0		0	1		1		0	1

$0.268 - \frac{6}{9} * 0.267 - \frac{1}{9} * 0.5$
$= -0.057$

AIR TIME GAP

2	15		28	30		35	40	68	90
1	1		0	0		1	0	0	0

★ [21.5 split] $0.268 - 0.25*0 - 0.75*0.167 = 0.143$ ★
[32.5 split] $0.268 - 0.5*0.333 - 0.5*0.25 = -0.024$
[37.5 split] $0.268 - \frac{5}{8}*0.3 - \frac{3}{8}*0 = 0.085$

INFO GAIN

$= IMPURITY (NODE\ D)$
$- \frac{N_{LEFT}}{N} \times IMPURITY (NODE\ D_{LEFT})$
$- \frac{N_{RIGHT}}{N} \times IMPURITY (NODE\ D_{RIGHT})$

$\Rightarrow$ AIR TIME GAP = ROOT

LEFT BRANCH
(ALL ATI GAP < 21.5)

0		1
1		1

OUTCOME IS HOMOGENOUS. NO MORE SPLITS...

RT BRANCH
(ALL ATI GAP ≥ 21.5)

0	0 0	0	0		1
0	0	0	0	1	0

$0.147 - \frac{1}{6}*0 - \frac{5}{6}*0.2 = 0$

(NO MORE SPLITS PAST THIS BECAUSE WE RAN OUT OF DIMENSIONS)

$\Rightarrow$ TREE =

idx	lst	rt	Split	Value
0	1	2	21.5	null
1	null	null	null	1
2	3	4	0.5	null
3	null	null	null	0
4	null	null	null	0

=

AIR_TIME_GAP < 21.5? ☜ ROOT

yes — DEP. DEL .15 = 1

no — PREV_DEP_DEL < 0.5?

yes — DEP_DEL .15 = 0

no — DEP_DEL_15 = 0