SIMULATE DT TRAINING

	STATE	CONTRACT	SEX	CALLS	CHURN	PRED.	
1	IT	C	F	>10	Y	Y	TP
2	GE	T	M	≤10	N	N	TN
3	IT	T	M	>10	Y	Y	TP
4	GE	Y	F	≤10	N	Y	FP
5	GE	T	F	>10	N	N	TN
6	GE	C	M	>10	Y	Y	TP
7	GE	C	M	≤10	Y	Y	TP
8	GE	Y	F	≤10	Y	Y	TP
9	GE	Y	M	>10	N	N	TN
10	GE	T	F	≤10	N	N	TN
11	GE	Y	F	>10	Y	Y	TP
12	IT	C	F	>10	N	Y	FP

CONSIDER THE MISSCLASSIFICATION ERROR $\boxed{ME = 1 - \text{max prob.}}$

ROOT $Y = 6$ $ME = \frac{6}{12}$

$N = 6$

$TP = 6$
$TN = 4$
$FP = 2$

STATE

	IT	GE
Y	2	4
N	1	5

$ME(IT) = 1/3$

$ME(GE) = 4/9$

$E = \frac{1}{3} \cdot \frac{3}{12} + \frac{4}{9} \cdot \frac{9}{12} = \frac{5}{12}$

CONTRACT

	C	T	Y
Y	3	1	2
N	1	3	2

$ME(C) = 1/4$

$ME(T) = 1/4$

$ME(Y) = 2/4$

$E = \frac{1}{4} \cdot \frac{4}{12} + \frac{1}{4} \cdot \frac{4}{12} + \frac{2}{4} \cdot \frac{4}{12} = \boxed{4/12}$

	CT	Y
Y	4	2
N	4	2

$ME(CT) = 4/8$

$ME(Y) = 2/4$

$E = \frac{4}{8} \cdot \frac{8}{12} + \frac{2}{4} \cdot \frac{4}{12} = \frac{6}{12}$

	C	TY
Y	3	3
N	1	5

$ME(C) = 1/4$

$ME(TY) = 3/8$

$E = \frac{1}{4} \cdot \frac{4}{12} + \frac{3}{8} \cdot \frac{8}{12} = \frac{4}{12}$

	CY	T
Y	5	1
N	3	3

$ME(CY) = 3/8$

$ME(T) = 1/4$

$E = \frac{3}{8} \cdot \frac{8}{12} + \frac{1}{4} \cdot \frac{4}{12} = \frac{4}{12}$

SEX

	F	M
Y	3	3
N	4	2

$ME(F) = 3/7$

$ME(M) = 2/5$

$E = \frac{3}{7} \cdot \frac{7}{12} + \frac{2}{5} \cdot \frac{5}{12} = \frac{5}{12}$

CALLS

	≤10	>10
Y	2	4
N	3	3

$ME(\leq 10) = 2/5$

$ME(>10) = 3/7$

$E = \frac{2}{5} \cdot \frac{5}{12} + \frac{3}{7} \cdot \frac{7}{12} = \frac{5}{12}$

SELECT CONTRACT = C

$Y = 3 \quad N = 1 \quad ME = 1/4$

STATE

	IT	GE
Y	1	2
N	1	0

$ME(IT) = 1/2$

$ME(GE) = \cancel{0}$

$E = \frac{1}{2} \cdot \frac{2}{4} = \frac{1}{4}$

SEX

	F	M
Y	1	2
N	1	0

$ME(F) = 1/2$ $ME(M) = \emptyset$ $E = \frac{1}{2} \cdot \frac{2}{4} = \frac{1}{4}$

CALLS

	≤ 10	> 10
Y	1	2
N	0	1

$ME(\leq 10) = 0$ $ME(>10) = \frac{1}{3}$ $E = \frac{1}{3} \cdot \frac{3}{4} = \frac{1}{4}$

NO IMPROVEMENT $\Rightarrow$ LEAF

SELECT CONTRACT = T $Y = 1$ $N = 3$ $E = 1/3$

STATE

	IT	GE
Y	1	0
N	0	3

$ME(IT) = 0$ $E = \emptyset$ $ME(GE) = 0$

WE CAN STOP WITH STATE BECAUSE WE HAVE REACHED THE LOWEST ERROR FOR THE ME SO MAX GAIN

SELECT CONTRACT = Y

Y=2 N=2 E = 2/4

STATE

	IT	GE
Y	0	2
N	0	2

ME(IT) = 0 ME(GE) = 2/4 $E = \frac{2}{4} \cdot \frac{4}{4} = \frac{2}{4}$

SEX

	F	M
Y	2	0
N	1	1

ME(F) = 1/3 ME(M) = 0 $E = \frac{1}{3} \cdot \frac{3}{4} = \frac{1}{4}$

CALLS

	≤10	>10
Y	1	1
N	1	1

ME(≤10) = 1/2 ME(>10) = 1/2 $E = \frac{1}{2} \cdot \frac{2}{4} + \frac{1}{2} \cdot \frac{2}{4} = \frac{2}{4}$

SELECT ~~SEX=F~~ Y=2 N=1 E = 1/3

STATE

	IT	GE
Y	0	2
N	0	1

ME(IT) = 0 ME(GE) = 1/3 $E = \frac{1}{3} \cdot \frac{3}{3} = \frac{1}{3}$

CALLS

	≤10	>10
Y	1	1
N	1	0

ME(≤10) = 1/2 ME(>10) = 0 $E = \frac{1}{2} \cdot \frac{2}{3} = \frac{1}{3}$

NO IMPROVEMENTS