PAGE
DATE

Assignment 3 _ Students dataset

We have 14 record → 8 A, 6 Not A

Total entropy (S) = $-\frac{6}{14}\log_2(\frac{6}{14}) - \frac{8}{14}\log_2(\frac{8}{14})$
= 0.985

* Early registration:

ER → 0 → 4+ 4- , ER → 1 → 4+ 2-

$H(S|ER=0) = -\frac{4}{8}\log_2\frac{4}{8} - \frac{4}{8}\log_2\frac{4}{8} = \boxed{1}$

$H(S|ER=1) = -\frac{4}{6}\log_2\frac{4}{6} - \frac{2}{6}\log_2\frac{2}{6} = \boxed{0.91}$

$Gain_{ER} = 0.985 - 1\times\frac{8}{14} - 0.91\times\frac{6}{14} = \boxed{0.0235}$

* Finished Homework

FH → 0 → 3+ 4- , FH → 1 → 5+ 2-

$H(S|FH=0) = -\frac{3}{7}\log_2\frac{3}{7} - \frac{4}{7}\log_2\frac{4}{7} = \boxed{0.985}$

$H(S|FH=1) = -\frac{5}{7}\log_2\frac{5}{7} - \frac{2}{7}\log_2\frac{2}{7} = \boxed{0.8631}$

$Gain_{FH} = 0.985 - 0.985\times\frac{7}{14} - 0.863\times\frac{7}{14}$
$= \boxed{0.061}$

* Senior

Senior → 0 ↗ 3+ , Senior → 1 ↗ 3+

$H(S|Senior=0) = \boxed{1}$

$H(S|Senior=1) = -5/8 \log_2 5/8 - 3/8 \log_2 3/8 = \boxed{0.954}$

$G_{senior} = 0.985 - 1 \times 6/14 - 0.954 \times 8/14 = \boxed{0.0113}$

* Liked coffee

LC → 0 ↗ 5+ , LC → 1 ↗ 3+

$H(S|LC=0) = 1$

$H(S|LC=1) = -3/4 \log_2 3/4 - 1/4 \log 1/4$
$= \boxed{0.811}$

$Gain_{LC} = 0.985 - 1 \times 10/14 - 0.811 \times 4/14$
$= \boxed{0.0598}$

* Liked last homework

LLH → 0 ↗ 3+ , LLH → 1 ↗ 5+

$H(S|LLH=0) = -3/5 \log_2 3/5 - 2/5 \log_2 2/5 = 0.970$

$H(S|LLH=1) = -5/9 \log_2 5/9 - 4/9 \log_2 4/9 = 0.991$

$G_{LLH} = 0.985 - 5/14 \times 0.970 - 9/14 \times 0.991 = \boxed{0.0015}$

As we see that finished last homework has the highest information gain. then it will be our tree root node.

We start calculating again.

Finished Homework

① Then get the gain of the rest

* ER → 0 (4): 2+, 2-

ER → 1 (5): 3+, 0-

$G_{ER} = 0.413$

* $G_{senior} = 0.167$

$G_{LC} = 0.34$

$G_{LLH} = 0.167$

⓪

* ER → 0 (4): 2+, 2-

ER → 1 (3): 1+, 2-

$G_{ER} = 0.02$

* Senior → 0 (4): 1+, 3-

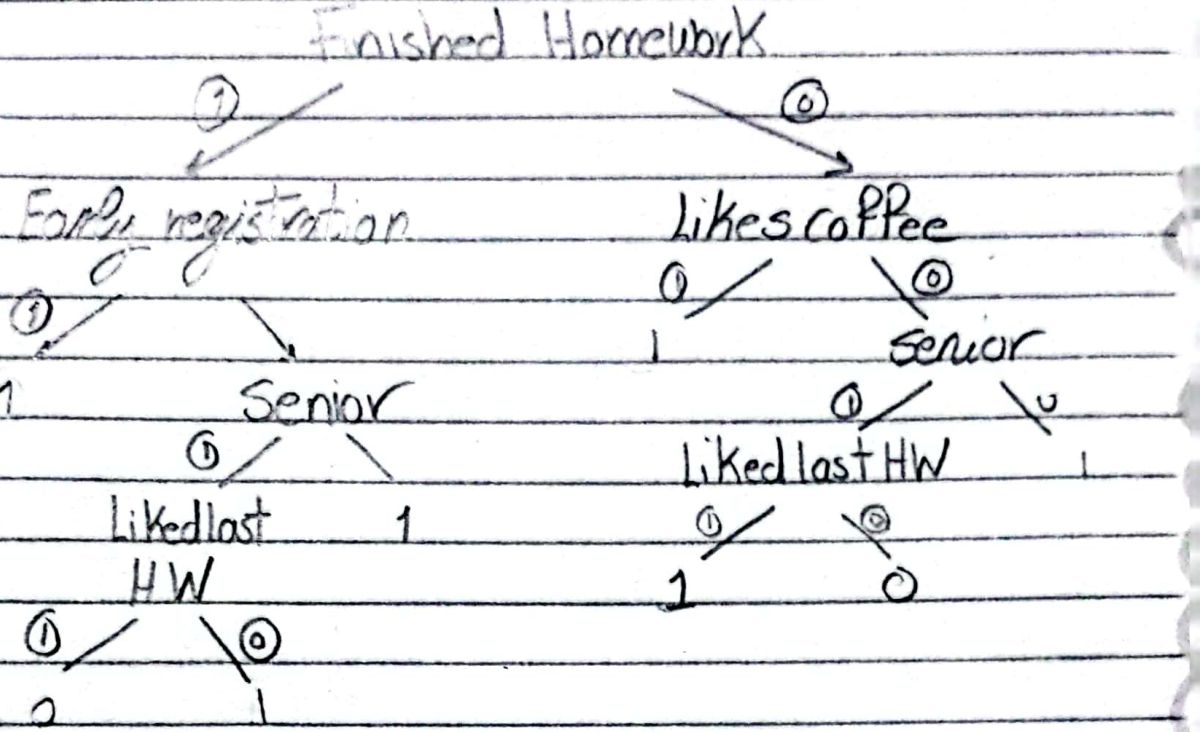
Senior → 1 (3): 2+, 1-

$G_{senior} = 0.128$

$G_{LC} = 0.47$

$G_{LLH} = 0.02$

Then we will take
early registration → ①
Liked Coffee → 0


Finished Homework
Early registration
Likes coffee
Senior
Liked last HW
Senior
Liked last HW