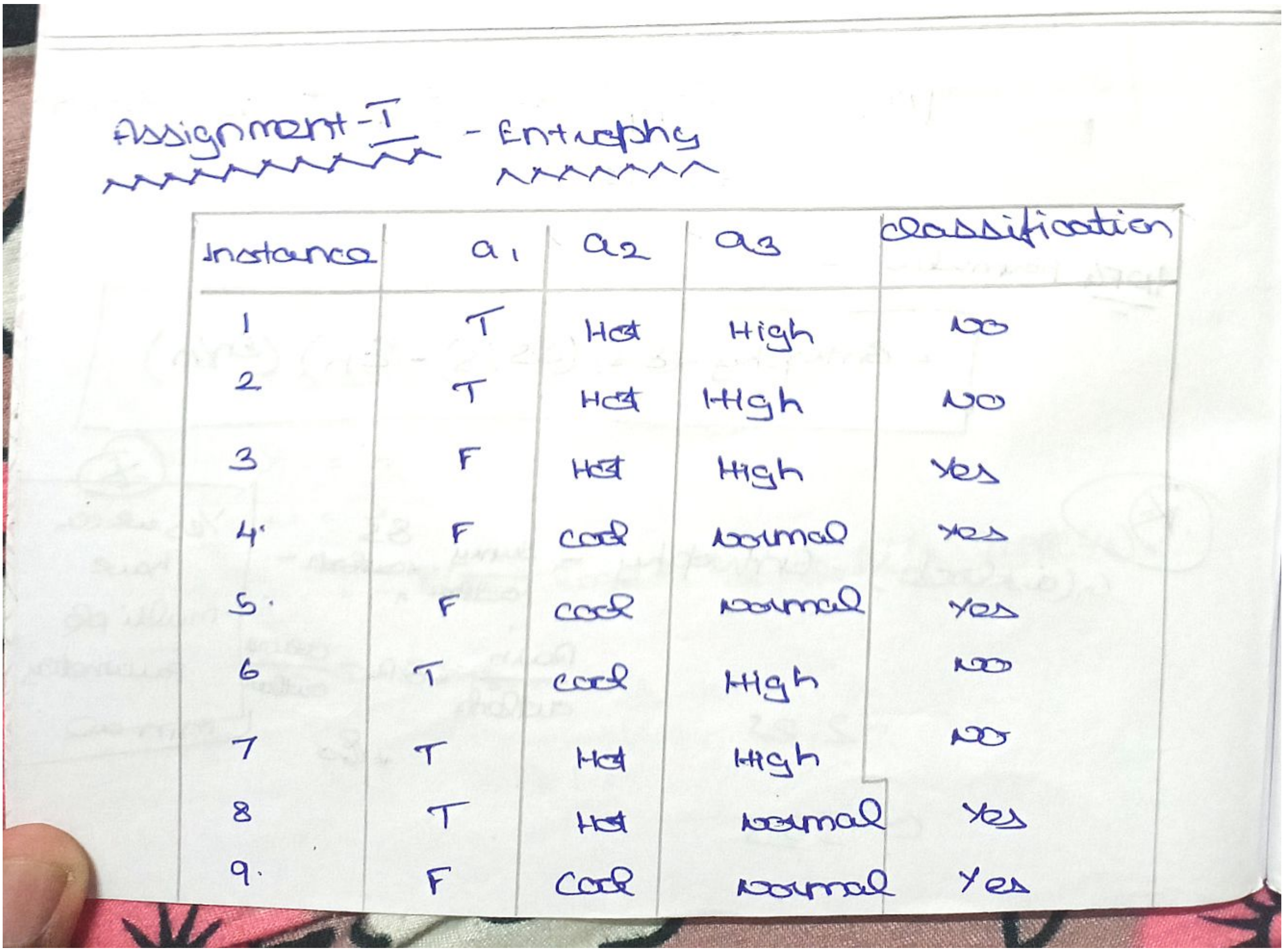


Assignment-I - Entrophy

Instance	a1	a2	a3	classification
1	T	Hot	High	NO
2	T	Hot	High	NO
3	F	Hot	High	Yes
4.	F	cool	Normal	Yes
5.	F	cool	Normal	Yes
6	T	cool	High	NO
7	T	Hot	High	NO
8	T	Hot	Normal	Yes
9.	F	Cool	Normal	Yes

10.	F	cool	High	Yes.

For calculating Enthophy:-

step 1:- overall Enthophy for the Data set

step 2: Find Individual Enthophy for each attribute

step 3: For that attribute calculate the Information Gain Formula.

step 4: Repeat 3 & 4 for all attributes.

step 5: Sum the Enthophy finded from all each attribute.

Calculation:-

1) overall Enthopy:-

Given

No. of DS = 10

Positive = $6/10$

Negative = $4/10$

$= -1.322$

$4/10 = 0.4$

$6/10 = -0.6$

$= ~~-1.322~~$

$= -0.737$

Enthophy Formula:-

$$\boxed{E(S) = -P_S \operatorname{Log}_2 P_S - N_S \operatorname{Log}_2 N_S}$$

$$E(S) = -6/10 \operatorname{Log}_2 6/10 - 4/10 \log_2 4/10$$

$$= 0.5288 + 0.4422$$

$$\boxed{E(S) = 0.971}$$

2) Entrophy & IGF (a1)

values - True, False (2)

True

Total - 5

T_s - 4/5

T_N - 1/5

4/5 = ·08

L = -0.3219

1/5 = ·32

L

= -2.322

False

Total - 5

F_s - 5/5

F_N - 0

(4) 000 are Yes

$\boxed{E(F) = 0}$

$E(T) = -\frac{4}{5}\log_2\frac{4}{5} - \frac{1}{5}\log_2\frac{1}{5}$

$= 0.25752 + 0.4644$

$\boxed{E(T) = 0.72192}$

INFORMATION GAIN FORMULA FOR a1 :-

$$\boxed{G(a_1) = \text{Entropy} - \text{True}/a_1 * E(T) - \text{False}/a_1 * E(F)}$$

$G(a_1) = 0.971 - 5/10 * 0.72192 - 5/10 * 0$

$= 0.971 - 0.36096 - 0$

$\boxed{G(a_1) = 0.61}$

3) Entrophy & IGF (a2) :-

values - Hot , cool

Hot

$3/5 = 0.6$

Total - 5

$L = 0.73$

$H_S - 3/5$

$2/5 = 0.4$

$H_N - 2/5$

$L = -1.322$

$E(H) = -3/5 \log_2 3/5 - 2/5 \log_2 2/5$

$= 0.4422 + 0.5288$

$= 0.971$

cool

4/5

Total - 5

$C_S - 4/5$

1/5

$C_N - 1/5$

$E(C) = -4/5 \log_2 4/5 - 1/5 \log_2 1/5$

$= 0$

$= 0.72192$

IFG Fa a2:-

$$G(a_2) = 0.971 - 5/10 * 0.971 - 5/10 * 0.72192$$

$= 0.971 - (0.5 * 0.971) - (0.5 * 0.72192)$

$- 0.4855 \quad - 0.36096$

$= 0.971$

$= 0.971 - 0.84646$

$$\boxed{G(a_2) = 0.124}$$

3) $\underline{a_3}$:-

value = 2

$\underline{H}$

T = 6

$H_S = 2/6$

$H_N = 4/6$

$2/6 = 0.33$

$L = -1.585$

$4/6 = 0.67$

$L = -0.585$

$= -2/6 \log_2 2/6 - 4/6 \log_2 4/6$

$= 0.52305 + 0.39195$

$= 0.915$

$\underline{N}$

T = 4

$N_S = 4/4$

$N_N = 0$

= 0

= 0

IGF :-

$= 0.917 - 6/10 * 0.915 - 0$

$= 0.917 - 0.549$

$\boxed{G(a_3) = 0.368}$

$a_1 = 0.61$

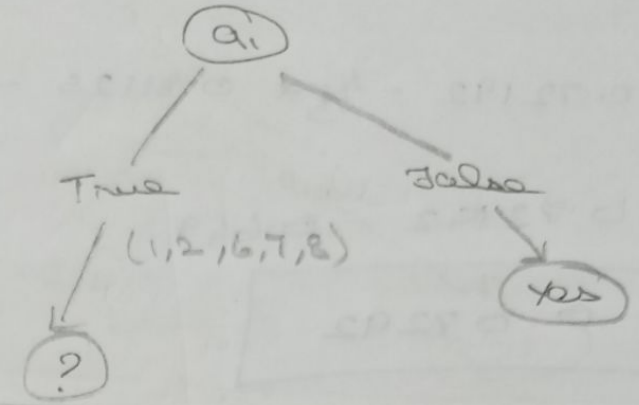
$a_2 = 0.124$

$a_3 = 0.368$

∴ From the calculated IG we can able to find the root node

Decision tree

Root node



(*) From False we get a conclusion, But in true we have to again learn the dataset belongs to true.

	a_1	a_2	a_3	classification
1	T	Hot	High	NO
2	T	Hot	High	NO
6	T	cool	High	NO
7	T	Hot	High	NO
8	T	Hot	Normal	Yes

By using this Dataset again we have to calculate next i[illegible] [illegible] attributes

$1/5 = 0.2$, $L = -2.322$, $4/5 = 0.8$, $L = -0.321$

$$E(S) = -1/5 \log_2 1/5 - 4/5 \log_2 4/5$$
$$= 0.4644 + 0.25752$$

$$E(S) = 0.72192$$

a_2 :- Entrophy

values - Hot, cool

Hot

T - 4

H_S - 3/4

H_N - 1/4

$$E(H) = -3/4 \log_2 3/4 - 1/4 \log_2 1/4$$
$$= 0.31125 + 0.5$$
$$= 0.81125$$

cool

T - 1

C_S - 0

C_N - 1/1

$E(C) = 0$

$3/4 = 0.75$

$L = 0.415$

$1/4 = 0.25$

$L = 2$

IGF a2

$= 0.72192 - 4/5 * 0.81125 - 0$

$= 0.72192 - 0.649$

$\boxed{G(a_2) = 0.07292}$

a_3 :-

Entrophy :-

values - High, Normal

High - 4

HS - 0

HN - 4/4

$\boxed{E(H) = 0}$

NO

Normal - 1

NS - 1/1

NN - 0

$\boxed{E(N) = 0}$

Yes

IGF :-

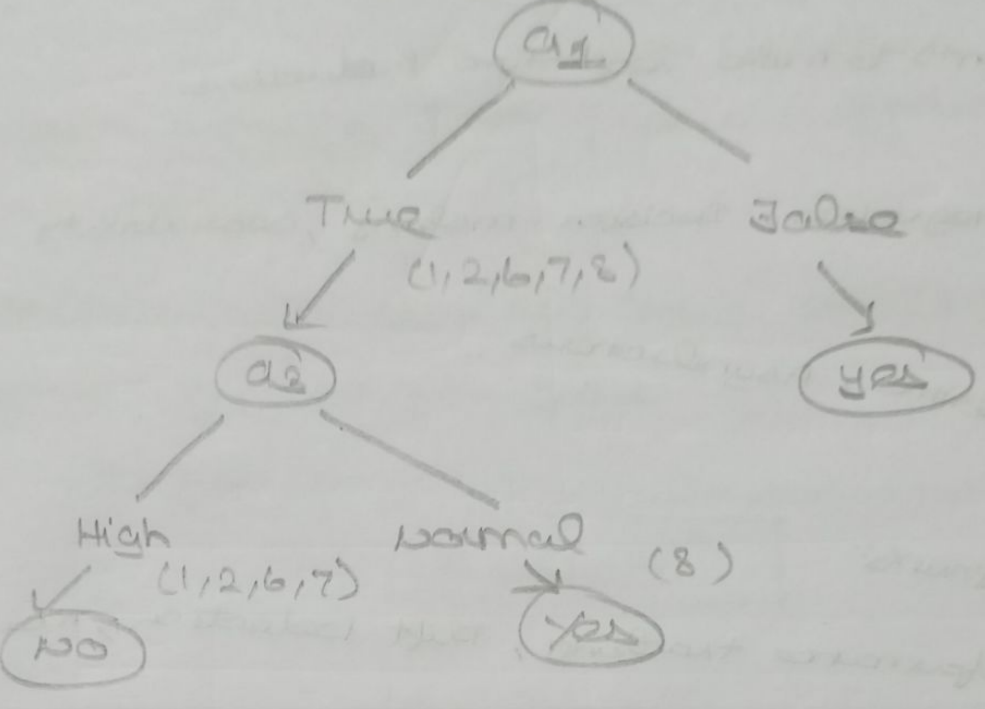
$= 0.72192 - 0 - 0$

$\boxed{G(a_3) = 0.72192}$

$a_2 = 0.07292$

$a_3 = 0.72192$

Decision Tree


a1
True
False
(1,2,6,7,8)
a2
yes
High
normal
(1,2,6,7)
(8)
no
yes