For task 2, you are required to analytically build a decision tree from the data given below in the table. Note that the label (target variable) is the "Decision" column. The data is about whether to play tennis or not given the weather conditions. So when the tree is built, it should be able to give out a decision on whether to play tennis or not given the weather conditions.

The general process is as follows:

i. At the first split starting from the root, we choose the attribute that has the max gain.

ii. Then, we re-start the same process at each of the children nodes (if node not pure).

iii. If node is pure, i.e. the node contains only the examples from a single class, then we keep it as a terminal node

Outlook	Temp	Humidity	Wind	Decision
Sunny	Hot	High	Weak	No
Sunny	Hot	High	Strong	No
Overcast	Hot	High	Weak	Yes
Rain	Mild	High	Weak	Yes
Rain	Cool	Normal	Weak	Yes
Rain	Cool	Normal	Strong	No
Overcast	Cool	Normal	Strong	Yes
Sunny	Mild	High	Weak	No
Sunny	Cool	Normal	Weak	Yes
Rain	Mild	Normal	Weak	Yes
Sunny	Mild	Normal	Strong	Yes
Overcast	Mild	High	Strong	Yes
Overcast	Hot	Normal	Weak	Yes
Rain	Mild	High	Strong	No

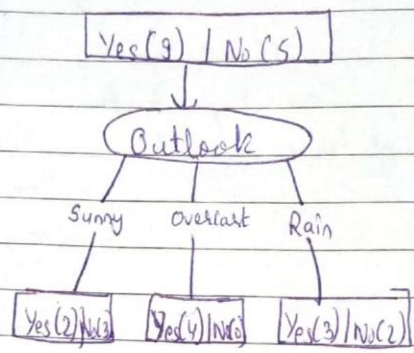
The following are the images of the above question's solution:

Lab Task

$\text{Entropy}(S) = -p_+ \log_2(p_+) - p_- \log_2(p_-)$

$\text{Gain}(S, A) = \text{Entropy}(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} \text{Entropy}(S_v)$

Outlook



$\text{Entropy}(9,5) = -p_+ \times \log_2(p_+) - p_- \times \log_2(p_-)$
$= -(9/14) \times \log_2(9/14) - (5/14) \times \log_2(5/14)$
$= 0.940$

Sunny:

$\text{Entropy}(2,3) = -p_+ \log_2(p_+) - p_- \times \log_2(p_-)$
$= -(2/5) + \log_2(2/5) - (3/5) \times \log_2(3/5)$
$= 0.970$

Overcast:

$\text{Entropy}(4,0) = -p_+ \times \log_2(p_+) - p_- \times \log_2(p_-)$
$= -(4/4) \times \log_2(4/4) - (0/4) \times \log_2(0/4)$
$= 0$

Rain:

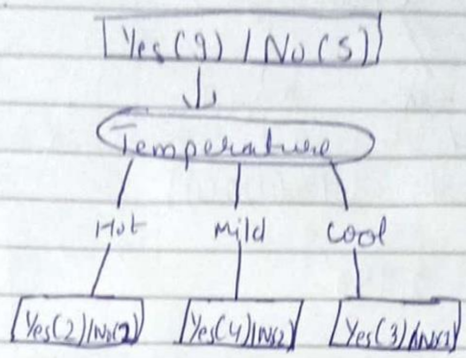
$\text{Entropy}(3,2) = -p_+ \times \log_2(p_+) - p_- \times \log_2(p_-)$
$= -(3/5) \times \log_2(3/5) - (2/5) \times \log_2(2/5)$
$= 0.971$

$p_+ (2/5)$ $p_- (3/5)$

~~Outlook~~ Gain(S, outlook) = $0.940 - (5/14) \times 0.970 - 0 - (5/14) \times 0.971$

$= 0.246$

Temperature.



Entropy (9,5) = ~~0.40~~ 0.940

Hot:

Entropy (2,2) $= -p_+ \times \log_2(p_+) - p_- \times \log_2(p_-)$

$= -(2/4) \times \log_2(2/4) - (2/4) \times \log_2(2/4)$

$= 1$

Mild:

Entropy (4,2) $= -p_+ \times \log_2(p_+) - p_- \times \log_2(p_-)$

$= -(4/6) \times \log_2(4/6) - (2/6) \times \log_2(2/6)$

$= 0.918$

Cool:

Entropy (3,1) $= -p_+ \times \log_2(p_+) - p_- \times \log_2(p_-)$

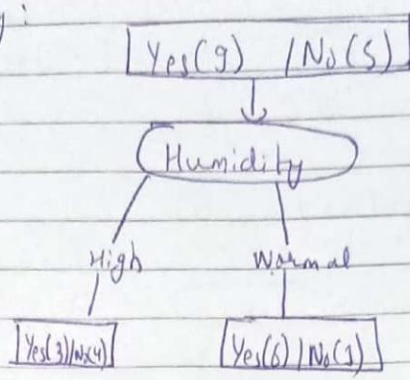
$= -(3/4) \times \log_2(3/4) - (1/4) \times \log_2(1/4)$

$= 0.811$

Gain(S, Temperature) $= 0.940 - (4/14) \times 1 - (6/14) \times 0.918 - (4/14) \times 0.811$

$= 0.029$

Humidity:



Entropy (9, 5) = 0.940

High:

Entropy(3, 4) = $-(3/7) \times \log_2(3/7) - (4/7) \times \log_2(4/7)$

= 0.985

Normal:

Entropy(6, 1) = $-(6/7) \times \log_2(6/7) - (1/7) \times \log_2(1/7)$

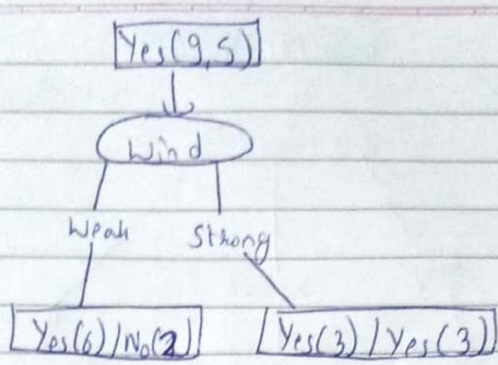
= 0.591

~~Humidity:~~

~~Gain (S, Entropy (S~~

Gain (S, Humidity) = $0.940 - (7/14) \times 0.985) - (7/14) \times 0.591$

= 0.152



Wind:

~~Entropy(9,5) = -(6/8) X log_2(6/8) - (2/8) X log_2(6/8)~~

~~= 0.8~~

Wind:

Entropy(9,5) = 0.940

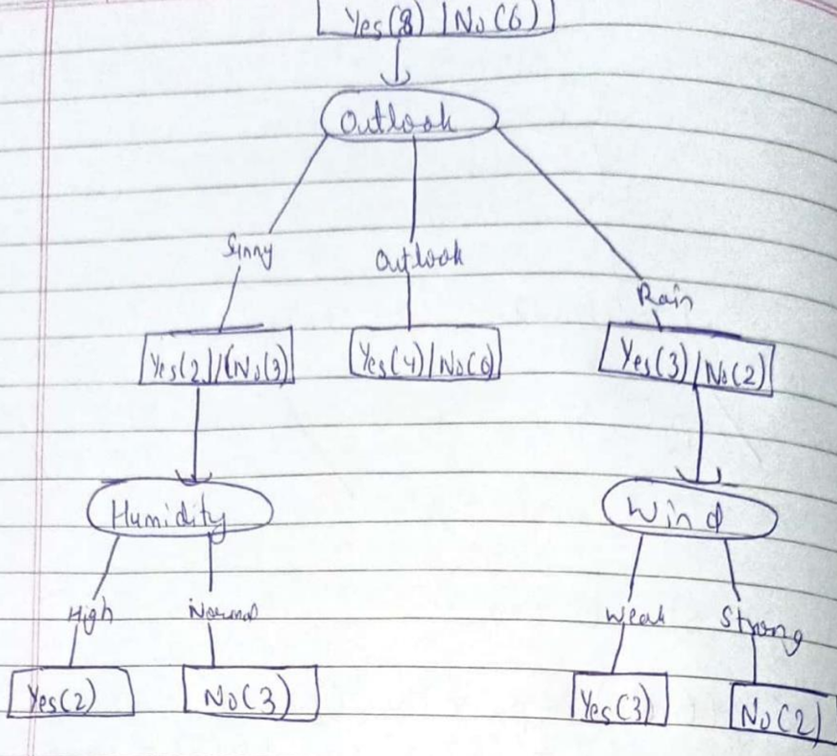
Weak:

$$\text{Entropy}(6,2) = -p_+ \times \log_2(p_+) - p_- \times \log_2(p_-)$$
$$= -(6/8) \times \log_2(6/8) - (2/8) \times \log_2(6/8)$$
$$= 0.811$$

$$\text{Entropy}(3,3) = -p_+ \times \log_2(p_+) - p_- \times \log_2(p_-)$$
$$= -(3/6) \times \log_2(3/6) - (3/6) \times \log_2(3/6)$$
$$= 1$$

$$\text{Gain}(S, \text{Wind}) = 0.940 - (8/14) \times 0.811 - (6/14) \times 1)$$
$$= 0.048$$

Feature	Information Gain
Outlook	0.246
Temperature	0.029
Humidity	0.152
Wind	0.048



final Decision tree