Decision Tree

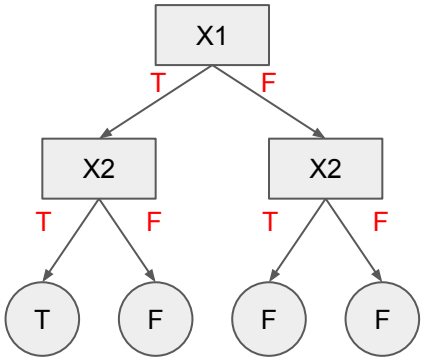
MACHINE LEARNING

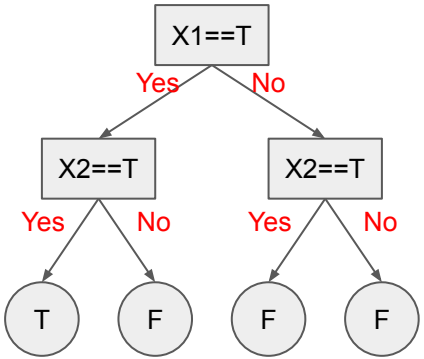
Pakarat Musikawan

Decision tree - 101

- Intuitive appeal for users
- Presentation Forms
  - “IF-Then” statement (decision rule)
    - IF … AND/OR … then ….
    - Partition sample of data
  - Graphical visualization (decision tree)
    - Work like a flowchart
    - Look like an upside down tree
    - Internal node and its branches
      - decision rules
    - Leaf node
      - a class label

X1	X2	Y
F	F	F
F	T	F
T	F	F
T	T	T



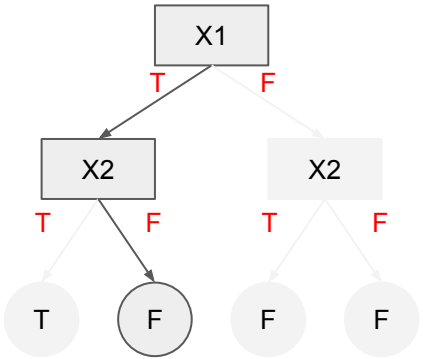


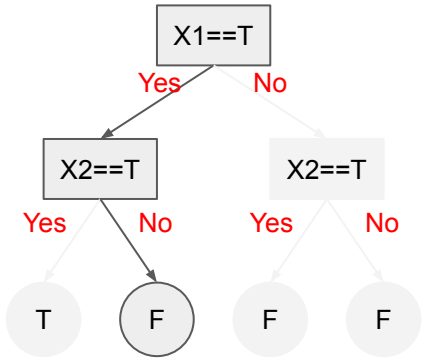
Decision tree - 101

X1	X2	Y
F	F	F
F	T	F
T	F	F
T	T	T

Decision tree - 101

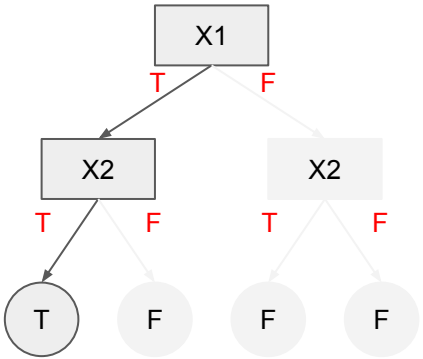
X1	X2	Y
F	F	F
F	T	F
T	F	F
T	T	T

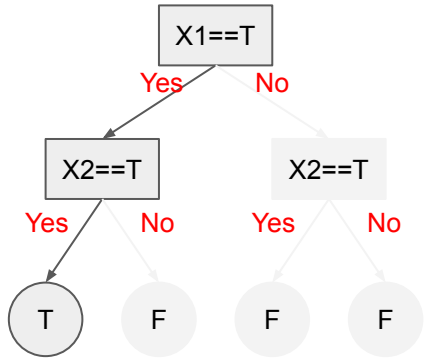




Decision tree - 101

X1	X2	Y
F	F	F
F	T	F
T	F	F
T	T	T





Sample dataset

- Rows denote labeled instances $\langle \boldsymbol{x}_i, y_i \rangle$
- Class label denotes whether a tennis game was played

$\langle \boldsymbol{x}_i, y_i \rangle$

Predictors				Response
Outlook	**Temperature**	**Humidity**	**Wind**	**Class**
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

Outlook = {Sunny, Overcast,Rain}

Temperature = {Hot,Mild,Cool}

Humidity = {High,Normal}

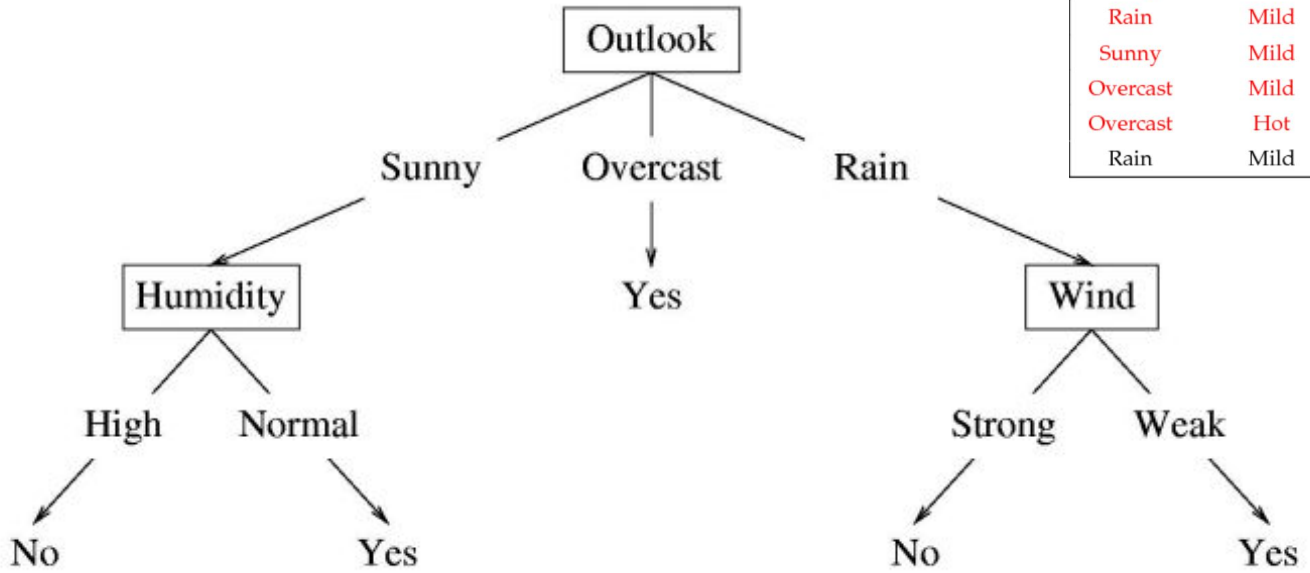
Wind = {Strong,Week}

Class = {Yes,No}

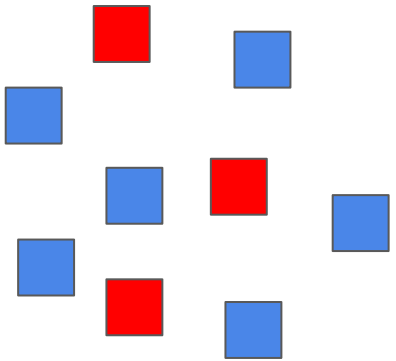
Decision tree

A trained decision tree for a mentioned dataset

Outlook	Temperature	Humidity	Wind	PlayTennis
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



Decision tree—ID3



$$Entropy = -\sum p(x)\log p(x)$$

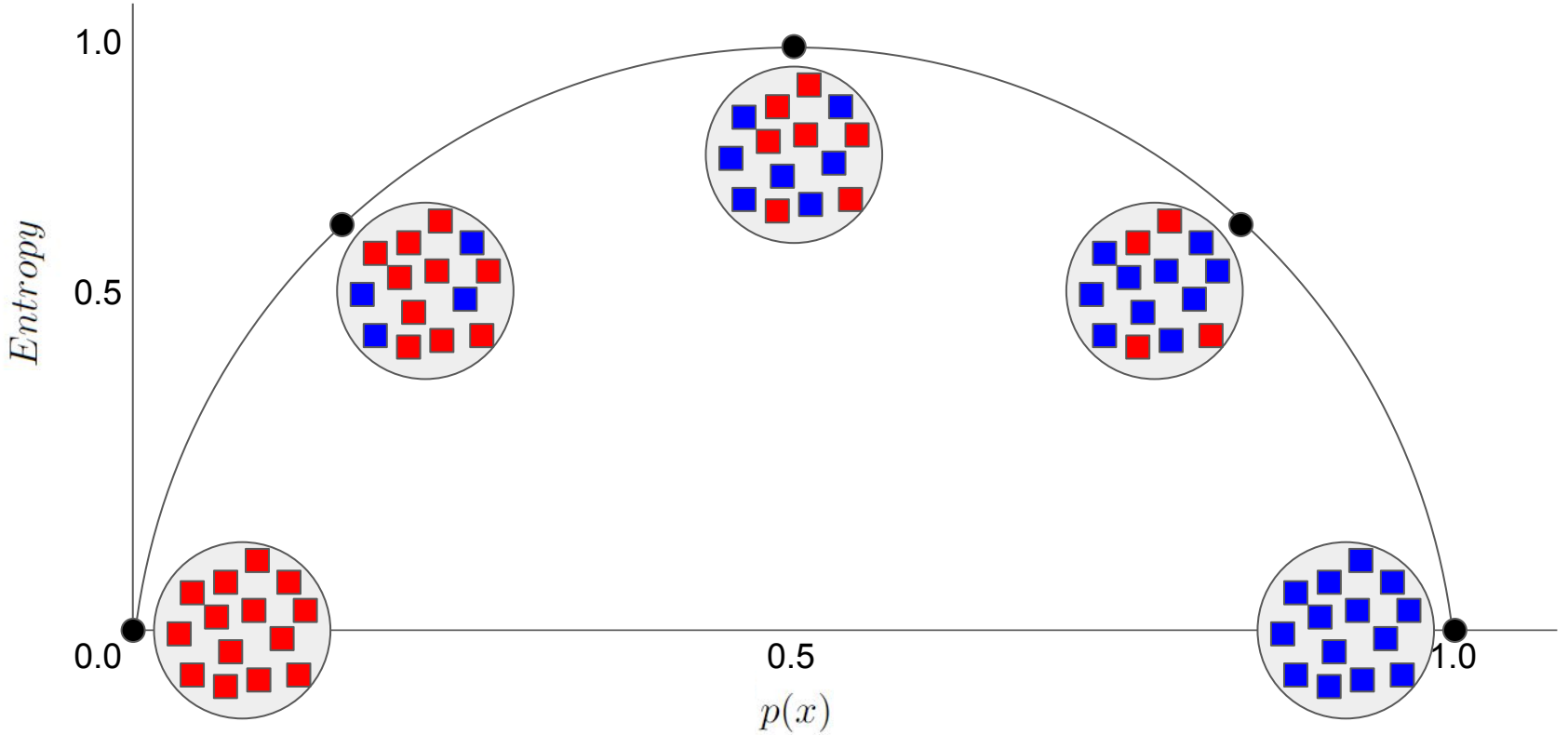
$$E(3,6) = -\left\{\left(\frac{3}{9}\log_2\frac{3}{9}\right) + \left(\frac{6}{9}\log_2\frac{6}{9}\right)\right\}$$

$$= -(-0.52 + -0.39)$$

$$= 0.91$$

Decision tree—ID3

$$Entropy = -\sum p(x) \log p(x)$$



Decision tree—ID3

Outlook	Temperature	Humidity	Wind	PlayTennis
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

PlayTennis (Y)	
Yes	No
9	5

$$E(Y) = -\sum p(x)\log p(x)$$

$$E(9,5) = -\left\{\left(\frac{9}{14}\log_2\frac{9}{14}\right) + \left(\frac{5}{14}\log_2\frac{5}{14}\right)\right\}$$

$$= -(-0.40 + -0.53)$$

$$= 0.93$$

Decision tree—ID3

Outlook	Temperature	Humidity	Wind	PlayTennis
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

Outlook (X)	Yes	No	Total
Sunny (S)	2	3	5
Overcast (O)	4	0	4
Rain (R)	3	2	5

$$
\begin{aligned}
E(Y|X) &= p(\mathrm{S})E(2,3) + p(\mathrm{O})E(4,0) + p(\mathrm{R})E(3,2) \\
&= \left(\frac{5}{14} \times 0.97\right) + \left(\frac{4}{14} \times 0\right) + \left(\frac{5}{14} \times 0.97\right) \\
&= 0.346 + 0 + 0.346 \\
&= 0.69 \\
IG(Y,X) &= E(Y) - E(Y|X) \\
&= 0.93 - 0.69 \\
&= 0.24
\end{aligned}
$$

Decision tree—ID3

Outlook	Temperature	Humidity	Wind	PlayTennis
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

Temperature (X)	Yes	No	Total
Hot (H)	2	2	4
Mild (M)	4	2	6
Cool (C)	3	1	4

$$
\begin{aligned}
E(Y|X) &= p(\mathrm{H})E(2,2) + p(\mathrm{M})E(4,2) + p(\mathrm{C})E(3,1) \\
&= \left(\frac{4}{14} \times 1\right) + \left(\frac{6}{14} \times 0.92\right) + \left(\frac{4}{14} \times 0.81\right) \\
&= 0.285 + 0.428 + 0.231 \\
&= 0.91 \\
IG(Y,X) &= E(Y) - E(Y|X) \\
&= 0.93 - 0.91 \\
&= 0.02
\end{aligned}
$$

Decision tree—ID3

Outlook	Temperature	Humidity	Wind	PlayTennis
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

Humidity (X)	Yes	No	Total
High (H)	3	4	7
Normal (N)	6	1	7

$$
\begin{aligned}
E(Y|X) &= p(\mathrm{H})E(3,4) + p(\mathrm{N})E(6,1) \\
&= \left(\frac{7}{14} \times 0.81\right) + \left(\frac{7}{14} \times 0.58\right) \\
&= 0.405 + 0.29 \\
&= 0.695 \\
IG(Y,X) &= E(Y) - E(Y|X) \\
&= 0.93 - 0.695 \\
&= 0.235
\end{aligned}
$$

Decision tree—ID3

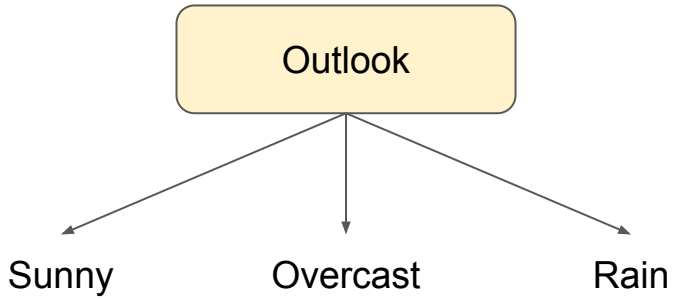
Outlook	Temperature	Humidity	Wind	PlayTennis
Sunny	Hot	High	Weak	No
Sunny	Hot	High	Strong	No
Overcast	Hot	High	Weak	Yes
Rain	Mild	High	Weak	Yes
Rain	Cool	Normal	Weak	Yes
Rain	Cool	Normal	Strong	No
Overcast	Cool	Normal	Strong	Yes
Sunny	Mild	High	Weak	No
Sunny	Cool	Normal	Weak	Yes
Rain	Mild	Normal	Weak	Yes
Sunny	Mild	Normal	Strong	Yes
Overcast	Mild	High	Strong	Yes
Overcast	Hot	Normal	Weak	Yes
Rain	Mild	High	Strong	No

Wind (X)	Yes	No	Total
Weak (W)	6	2	8
Strong (S)	3	3	6

$$
\begin{aligned}
E(Y|X) &= p(\mathrm{W})E(6,2) + p(\mathrm{S})E(3,3) \\
&= \left(\frac{8}{14} \times 0.81\right) + \left(\frac{6}{14} \times 1\right) \\
&= 0.46 + 0.42 \\
&= 0.88 \\
IG(Y,X) &= E(Y) - E(Y|X) \\
&= 0.93 - 0.88 \\
&= 0.05
\end{aligned}
$$

Decision tree—ID3

	IG(Y,X)
Outlook	0.24
Temperature	0.02
Humidity	0.235
Wind	0.05



Decision tree—ID3

Outlook	Temperature	Humidity	Wind	PlayTennis
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

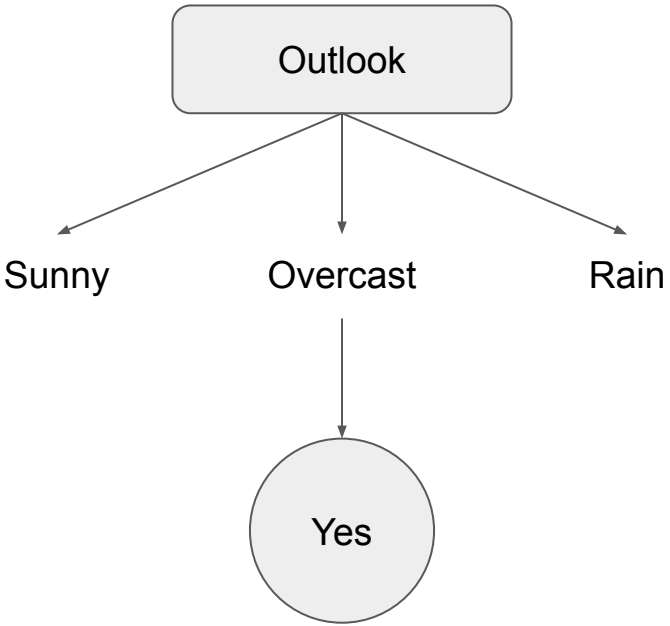
Outlook (Sunny)	
Yes	No
2	3

Outlook (Rain)	
Yes	No
3	2

Outlook (Overcast)	
Yes	No
4	0

Decision tree—ID3

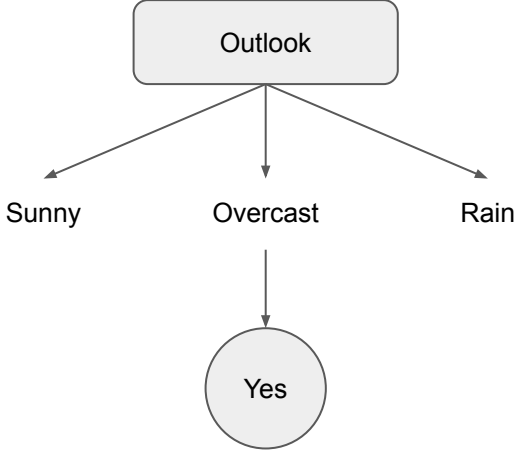
Outlook	Temperature	Humidity	Wind	PlayTennis
Sunny	Hot	High	Weak	No
Sunny	Hot	High	Strong	No
Overcast	Hot	High	Weak	Yes
Rain	Mild	High	Weak	Yes
Rain	Cool	Normal	Weak	Yes
Rain	Cool	Normal	Strong	No
Overcast	Cool	Normal	Strong	Yes
Sunny	Mild	High	Weak	No
Sunny	Cool	Normal	Weak	Yes
Rain	Mild	Normal	Weak	Yes
Sunny	Mild	Normal	Strong	Yes
Overcast	Mild	High	Strong	Yes
Overcast	Hot	Normal	Weak	Yes
Rain	Mild	High	Strong	No



Decision tree—ID3

Outlook	Temperature	Humidity	Wind	PlayTennis
Sunny	Hot	High	Weak	No
Sunny	Hot	High	Strong	No
Sunny	Mild	High	Weak	No
Sunny	Cool	Normal	Weak	Yes
Sunny	Mild	Normal	Strong	Yes

PlayTennis (Y) Outlook = Sunny	
Yes	No
2	3



$$E(Y) = -\sum p(x) \log p(x)$$

$$E(2,3) = -\left\{ \left(\frac{2}{5} \log_2 \frac{2}{5} \right) + \left(\frac{3}{5} \log_2 \frac{3}{5} \right) \right\}$$

$$= 0.97$$

Decision tree—ID3

Outlook	Temperature	Humidity	Wind	PlayTennis
Sunny	Hot	High	Weak	No
Sunny	Hot	High	Strong	No
Sunny	Mild	High	Weak	No
Sunny	Cool	Normal	Weak	Yes
Sunny	Mild	Normal	Strong	Yes

Temperature (X)	Yes	No	Total
Hot (H)	0	2	2
Mild (M)	1	1	2
Cool (C)	1	0	1

$$
\begin{aligned}
E(Y|X) &= p(\mathrm{H})E(0,2) + p(\mathrm{M})E(1,1) + p(\mathrm{C})E(1,0) \\
&= \left(\frac{2}{5} \times 0\right) + \left(\frac{2}{5} \times 1\right) + \left(\frac{1}{5} \times 0\right) \\
&= 0.4 \\
IG(Y,X) &= E(Y) - E(Y|X) \\
&= 0.57
\end{aligned}
$$

Decision tree—ID3

Outlook	Temperature	Humidity	Wind	PlayTennis
Sunny	Hot	High	Weak	No
Sunny	Hot	High	Strong	No
Sunny	Mild	High	Weak	No
Sunny	Cool	Normal	Weak	Yes
Sunny	Mild	Normal	Strong	Yes

Humidity (X)	Yes	No	Total
High (H)	0	3	3
Normal (N)	2	0	2

$$E(Y|X) = p(\mathrm{H})E(0,3) + p(\mathrm{N})E(0,2)$$

$$= \left(\frac{3}{5} \times 0\right) + \left(\frac{2}{5} \times 0\right)$$

$$= 0$$

$$IG(Y,X) = E(Y) - E(Y|X)$$

$$= 0.97$$

Decision tree—ID3

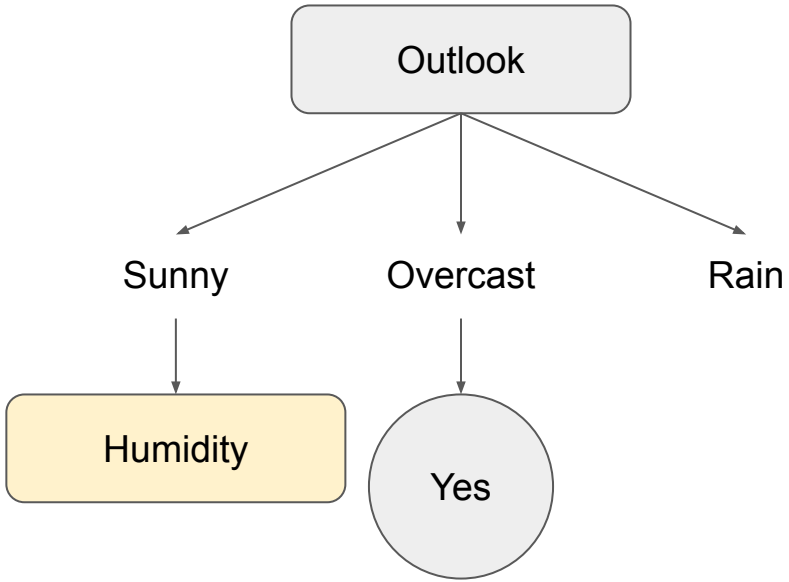
Outlook	Temperature	Humidity	Wind	PlayTennis
Sunny	Hot	High	Weak	No
Sunny	Hot	High	Strong	No
Sunny	Mild	High	Weak	No
Sunny	Cool	Normal	Weak	Yes
Sunny	Mild	Normal	Strong	Yes

Wind (X)	Yes	No	Total
Weak (W)	1	2	3
Strong (S)	1	1	2

$$
\begin{aligned}
E(Y|X) &= p(\mathrm{W})E(1,2) + p(\mathrm{S})E(1,1) \\
&= \left(\frac{3}{5} \times 0.92\right) + \left(\frac{2}{5} \times 1\right) \\
&= 0.952 \\
IG(Y,X) &= E(Y) - E(Y|X) \\
&= 0.018
\end{aligned}
$$

Decision tree—ID3

	IG(Y,X)
Temperature	0.57
Humidity	0.97
Wind	0.018

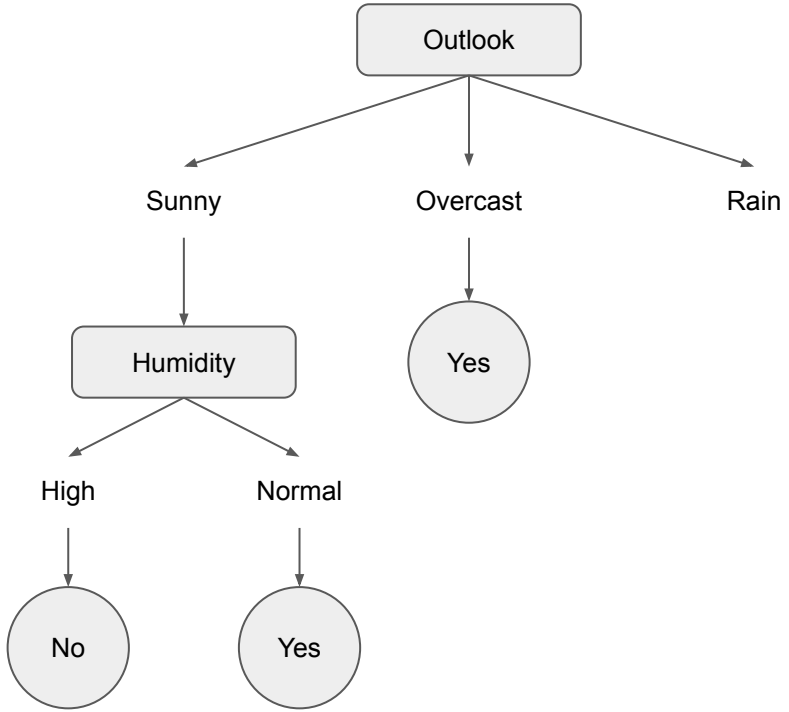


Decision tree—ID3

Outlook	Temperature	Humidity	Wind	PlayTennis
Sunny	Hot	High	Weak	No
Sunny	Hot	High	Strong	No
Sunny	Mild	High	Weak	No
Sunny	Cool	Normal	Weak	Yes
Sunny	Mild	Normal	Strong	Yes

Humidity (High)	
Yes	No
0	3

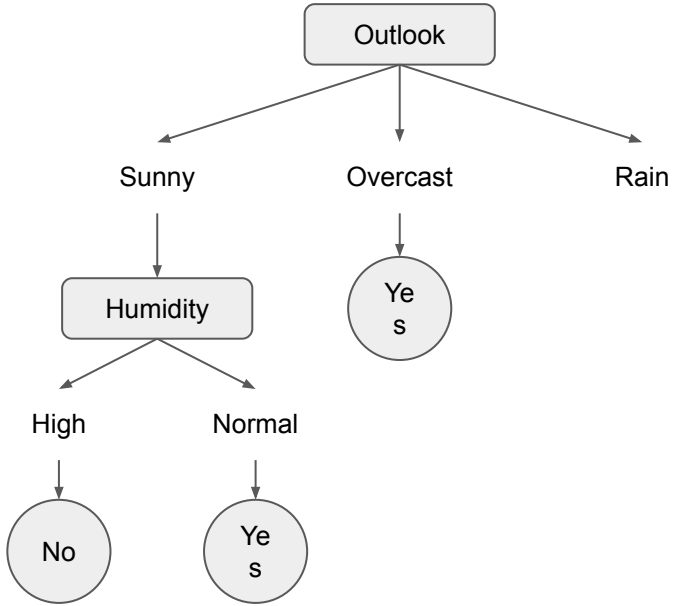
Humidity (Normal)	
Yes	No
2	0



Decision tree—ID3

Outlook	Temperature	Humidity	Wind	PlayTennis
Rain	Mild	High	Weak	Yes
Rain	Cool	Normal	Weak	Yes
Rain	Cool	Normal	Strong	No
Rain	Mild	Normal	Weak	Yes
Rain	Mild	High	Strong	No

PlayTennis (Y) Outlook = Rain	
Yes	No
3	2



$$E(Y) = -\sum p(x) \log p(x)$$

$$E(3,2) = -\left\{ \left(\frac{3}{5} \log_2 \frac{3}{5} \right) + \left(\frac{2}{5} \log_2 \frac{2}{5} \right) \right\}$$

$$= 0.97$$

Decision tree—ID3

Outlook	Temperature	Humidity	Wind	PlayTennis
Rain	Mild	High	Weak	Yes
Rain	Cool	Normal	Weak	Yes
Rain	Cool	Normal	Strong	No
Rain	Mild	Normal	Weak	Yes
Rain	Mild	High	Strong	No

Temperature (X)	Yes	No	Total
Hot (H)	0	0	0
Mild (M)	2	1	3
Cool (C)	1	1	2

$$E(Y|X) = p(\mathrm{H})E(0,0) + p(\mathrm{M})E(2,1) + p(\mathrm{C})E(1,1)$$

$$= \left(\frac{0}{5} \times 0\right) + \left(\frac{3}{5} \times 1\right) + \left(\frac{2}{5} \times 0\right)$$

$$= 0.6$$

$$IG(Y,X) = E(Y) - E(Y|X)$$

$$= 0.37$$

Decision tree—ID3

Outlook	Temperature	Humidity	Wind	PlayTennis
Rain	Mild	High	Weak	Yes
Rain	Cool	Normal	Weak	Yes
Rain	Cool	Normal	Strong	No
Rain	Mild	Normal	Weak	Yes
Rain	Mild	High	Strong	No

Humidity (X)	Yes	No	Total
High (H)	1	1	2
Normal (N)	2	1	3

$$
\begin{aligned}
E(Y|X) &= p(\mathrm{H})E(1,1) + p(\mathrm{N})E(2,1) \\
&= \left(\frac{2}{5} \times 1\right) + \left(\frac{3}{5} \times 0.92\right) \\
&= 0.952 \\
IG(Y,X) &= E(Y) - E(Y|X) \\
&= 0.018
\end{aligned}
$$

Decision tree—ID3

Outlook	Temperature	Humidity	Wind	PlayTennis
Rain	Mild	High	Weak	Yes
Rain	Cool	Normal	Weak	Yes
Rain	Cool	Normal	Strong	No
Rain	Mild	Normal	Weak	Yes
Rain	Mild	High	Strong	No

Wind (X)	Yes	No	Total
Weak (W)	3	0	3
Strong (S)	0	2	2

$$E(Y|X) = p(\mathrm{W})E(3,0) + p(\mathrm{S})E(0,2)$$

$$= \left(\frac{3}{5} \times 0\right) + \left(\frac{2}{5} \times 0\right)$$

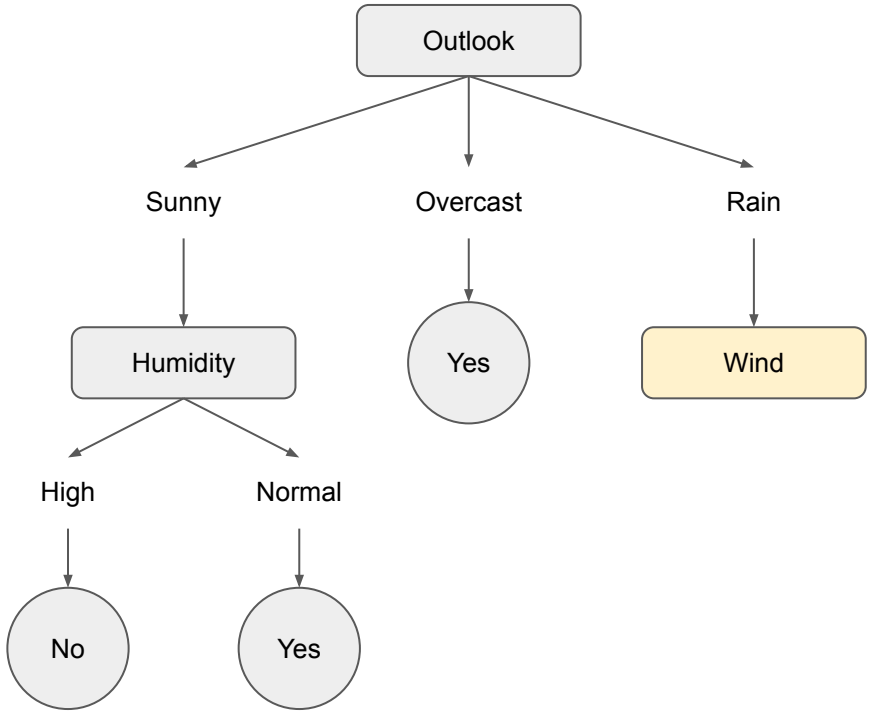
$$= 0$$

$$IG(Y,X) = E(Y) - E(Y|X)$$

$$= 0.97$$

Decision tree—ID3

	IG(Y,X)
Temperature	0.37
Humidity	0.018
Wind	0.97

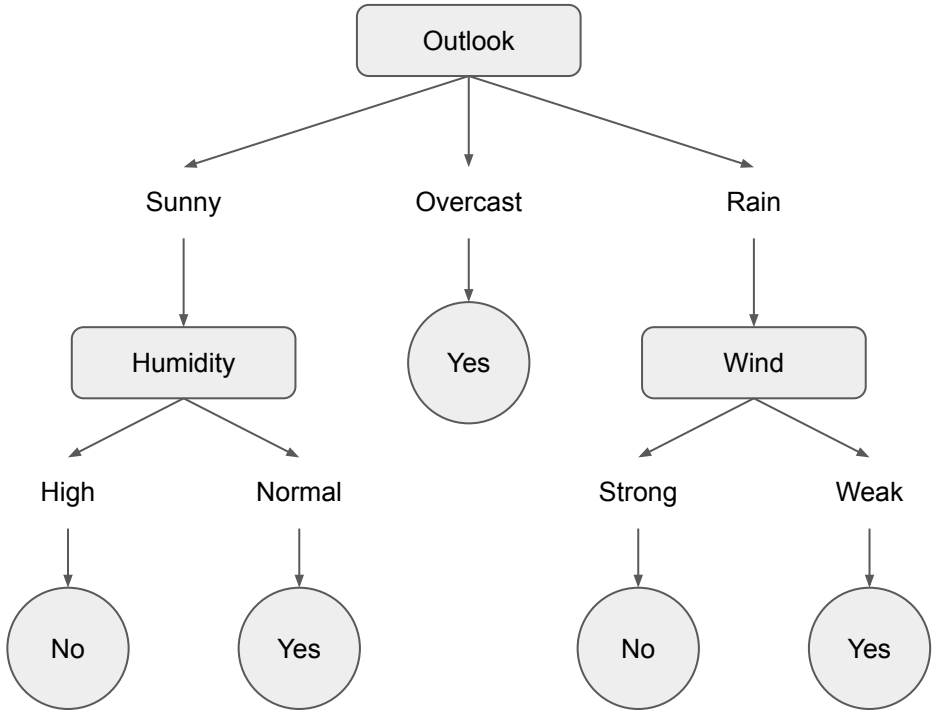


Decision tree—ID3

Outlook	Temperature	Humidity	Wind	PlayTennis
Rain	Mild	High	Weak	Yes
Rain	Cool	Normal	Weak	Yes
Rain	Cool	Normal	Strong	No
Rain	Mild	Normal	Weak	Yes
Rain	Mild	High	Strong	No

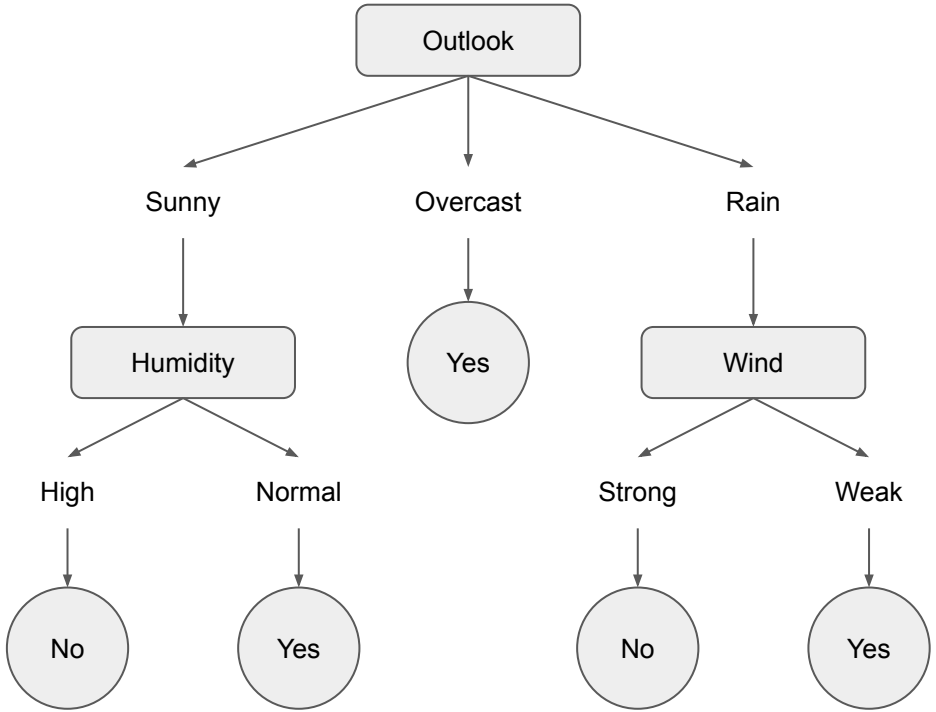
Wind (Weak)	
Yes	No
3	0

Wind (Strong)	
Yes	No
0	2



Decision tree—ID3

Outlook	Temperature	Humidity	Wind	PlayTennis
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



Workshop

ให้ทำการสร้าง Decision tree ด้วยขั้นตอนวิธี ID3 โดยใช้ข้อมูลที่กำหนดให้

X1	X2	X3	Y
False	False	False	False
False	False	True	False
False	True	False	False
False	True	True	True
True	False	False	False
True	False	True	True
True	True	False	True
True	True	True	True