

Online Assignment 3

Course code: CAP446

Course Title: Datawarehousing and Data mining

NAME : Jayshri Lal Pandit

Roll No: RD2112A103

Reg. No.: 12111670

Section : D21L2

① Given dataset :-

Day	Outlook	Temperature	Humidity	Wind	Play Golf?
1	Sunny	Hot	High	False	No
2	Sunny	Hot	High	True	No
3	Overcast	Hot	High	False	Yes
4	Rain	Mild	High	False	Yes
5	Rain	Cool	Normal	False	Yes
6	Rain	Cool	Normal	True	No
7	Overcast	Cool	Normal	True	Yes
8	Sunny	Mild	High	False	No
9	Sunny	Cool	Normal	False	Yes
10	Rain	Mild	Normal	False	Yes
11	Sunny	Mild	Normal	True	Yes
12	Overcast	Mild	High	True	Yes
13	Overcast	Hot	Normal	False	Yes
14	Rain	Mild	High	True	No

Given data sample :-

(Outlook = sunny, Temperature = Cool, Humidity = High, Wind = True)

* Step: ①

$$P(C_i) : P(\text{Play Golf} = \text{"Yes"}) = 9/14 = 0.643$$

$$P(\text{Play Golf} = \text{"No"}) = 5/14 = 0.357$$

Step ②: Compute $P(X/C_i)$ for each class

$$P(\text{Outlook} = \text{Sunny} | \text{Play Golf} = \text{"Yes"}) = 2/9 = 0.222$$

$$P(\text{Outlook} = \text{Sunny} | \text{Play Golf} = \text{"No"}) = 3/5 = 0.6$$

$$P(\text{Temperature} = \text{cool} \mid \text{play Golf} = \text{"yes"}) = 3/9 = 0.333$$

$$P(\text{Temperature} = \text{cool} \mid \text{play Golf} = \text{"No"}) = 1/5 = 0.2$$

$$P(\text{Humidity} = \text{High} \mid \text{play Golf} = \text{"yes"}) = 3/9 = 0.333$$

$$P(\text{Humidity} = \text{High} \mid \text{play Golf} = \text{"No"}) = 4/5 = 0.8$$

$$P(\text{Wind} = \text{True} \mid \text{play Golf} = \text{"yes"}) = 3/9 = 0.333$$

$$P(\text{Wind} = \text{True} \mid \text{play Golf} = \text{"No"}) = 3/5 = 0.6$$

Step 3:- $X = (\text{Outlook} = \text{Sunny}, \text{Temperature} = \text{cool}, \text{Humidity} = \text{High}, \text{Wind} = \text{True})$

$$P(X \mid C_i): P(X \mid \text{play Golf} = \text{"yes"})$$

$$P(X \mid \text{play Golf} = \text{"yes"}) = 0.222 \times 0.333 \times 0.333 \times 0.333 \\ = 0.0082$$

$$P(X \mid \text{play Golf} = \text{"No"}) = 0.6 \times 0.2 \times 0.8 \times 0.6 \\ = 0.057$$

Step 4:- $P(X \mid C_i) * P(C_i)$:

$$P(X \mid \text{play Golf} = \text{"yes"}) * P(\text{play Golf} = \text{"yes"}) \\ = 0.0082 * 0.643 = 0.0052$$

$$P(X \mid \text{play Golf} = \text{"No"}) * P(\text{play Golf} = \text{"No"}) \\ = 0.057 * 0.357 = 0.0203$$

Therefore, X belongs to Class ("play Golf = No")

(2.)

Decision Tree

Decision tree is a supervised learning algorithm which is used for both classification and regression. It uses a tree-like model of decisions, where each internal node denotes a test on an attribute, each branch represents an outcome of the test and each leaf node (terminal node) holds a class label. The topmost node in a tree is the root node.

It consists of a set of rules for dividing a large heterogeneous population into smaller groups with respect to the target label, the algorithm used for this tree construction is recursive partitioning. It follows top-down approach.

Decision Tree Terminology

Decision tree terminologies are given below :-

- **Root Node :-** It is the topmost node in decision tree, which represents entire population.
- **Splitting :-** It is a process, which

divide samples .

- Decision Node:- It is special type of node, which splits node into further sub nodes.
- Leaf / Terminal Node:- It is a last stage of node of decision tree, which holds a class label.
- Pruning:- It is opposite to splitting because it reduce size of tree.

Example of Decision Tree

Outlook	Temperature	Humidity	Wind	Play Golf
Sunny	Hot	High	False	No
Sunny	Hot	High	True	No
Overcast	Hot	High	False	Yes
Rain	Mild	High	False	Yes
Rain	Cool	Normal	False	Yes
Rain	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Sunny	Mild	High	False	No
Sunny	Cool	Normal	False	Yes
Rain	Mild	Normal	False	Yes
Sunny	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Rain	Mild	High	True	No

