Name: Atul Kumar Roll No.: RD2110B79
Reg. No: 12102801 Course Code: CAP-446
Instructor: Dr. Geeta Sharma CA: 03

Q1. Consider the given data set and perform Naive Bayesian classification. Mention each steps and calculation clearly.
(Outlook = Sunny, Temperature = Cool, Humidity = High, Wind = True)

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

Ans: Target Node = Play Golf

→ First we have to calculate the prior probability (Target node)

$$P(E) = \frac{\text{Total no. of fav. outcomes}}{\text{Total no. of outcomes}}$$

i) $P(\text{Play Golf} = \text{yes}) = \frac{9}{14} = 0.64$

ii) $P(\text{Play Golf} = \text{no}) = \frac{5}{14} = 0.36$

→ Now we have to calculate the conditional probability based on the given table for all the attributes like Outlook, Temperature, Humidity and Wind.

Outlook	Y	N
Sunny	2/9	3/5
Overcast	4/9	0
Rain	3/9	2/5

Temperature	Y	N
Hot	2/9	2/5
Mild	4/9	2/5
Cool	3/9	1/5

Humidity	Y	N
High	3/9	4/5
Normal	6/9	1/5

Wind	Y	N
True	3/9	3/5
False	6/9	2/5

→ From the given condition in the question we have to find the new instance for yes and no.

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

$$V_{NB} = \underset{v_j \in \{yes, no\}}{\operatorname{argmax}} P(v_j) \prod_i P(a_i | v_j)$$

$$= \underset{v_j \in \{yes, no\}}{\operatorname{argmax}} P(v_j) \; P(Outlook = Sunny | v_j) \; P(Temperature = Cool | v_j) \; P(Humidity = High | v_j) \; P(Wind = True | v_j)$$

Put the values in the above formula for yes and no

i) For yes

$$V_{NB}(Yes) = P(Yes) \; P(Sunny | Yes) \; P(Cool | Yes) \; P(High | Yes) \; P(True | Yes)$$

$$= 9/14 \times 2/9 \times 3/9 \times 3/9 \times 3/9$$

$$= 0.00526$$

ii) For no

$$V_{NB}(No) = P(No)\ P(Sunny|No)\ P(Cool|No)\ P(High|No)\ P(True|No)$$

$$= \frac{5}{14} \times \frac{3}{5} \times \frac{1}{5} \times \frac{4}{5} \times \frac{3}{5}$$

$$= 0.02057$$

→ Now calculate the probability for Yes and No if the probability of yes is greater than player can play golf and if no is greater than players can't play golf.

i) For Yes

$$V_{NB}(Yes) = \frac{V_{NB}(Yes)}{V_{NB}(Yes) + V_{NB}(No)}$$

$$= \frac{0.00526}{0.00526 + 0.02057}$$

$$= 0.2036$$

ii) For No

$$V_{N_B}(No) = \frac{V_{N_B}(No)}{V_{N_B}(Yes) + V_{N_B}(No)}$$

$$= \frac{0.02057}{0.00526 + 0.02057}$$

$$= 0.7963$$

$$\boxed{V_{N_B}(Yes) < V_{N_B}(No)}$$

Players can't play golf.

Q2 Briefly explain Decision tree and its terminologies by taking an example.

Ans: Decision Tree is a general, predictive modelling tool that has applications spanning a number of different areas. Decision are constructed by an algorithm approach that identifies ways to split a data set based on different conditions. The goal is to create a model that predicts the value of a target variable by learning simple decision rules inferred from the data features. It includes a root node, branches and leaf nodes. Each internal node denotes a test on an attribute, each branch denotes the outcome of a test, and each leaf node holds a class label. They are used in non-linear decision making with simple linear decision surface.

The benefits of having a decision tree are:-

i) It does not require any domain knowledge

ii) It is easy to comprehand

iii) The learning and classification steps of a decision tree are simple and fast.

→ There are two approaches to prune(cut) a tree

i) Pre-pruning: The tree is pruned by halting its construction early.

ii) Post-pruning: This approach removes a sub-tree from a fully constructed tree.

→ There are some terminologies of decision tree:

i) Instances: Attribute that define the input space.

ii) Attribute: A quantity describing an instan

iii) Hypothesis Class: Set of all the possible attributes.

iv) Root Node: The topmost node in the tree is the root node.

v) Non-leaf Node: Those node which have branches (children).

vi) Leaf Node: Those node which have no branches (children).

Eg: Draw a decision tree for the concept Play Badminton

Day	Weather	Temperature	Humidity	Play
1	Sunny	Hot	High	No
2	Cloudy	Hot	High	No
3	Sunny	Mild	Normal	Yes
4	Cloudy	Mild	High	No
5	Rainy	Mild	High	No
6	Rainy	Cool	Normal	~~No~~
7	Rainy	Mild	High	No
8	Sunny	Hot	High	No
9	Cloudy	Hot	Normal	Yes
10	Rainy	Mild	High	No

Observations of the last ten days.

Decision Tree

Root Node:- Weather

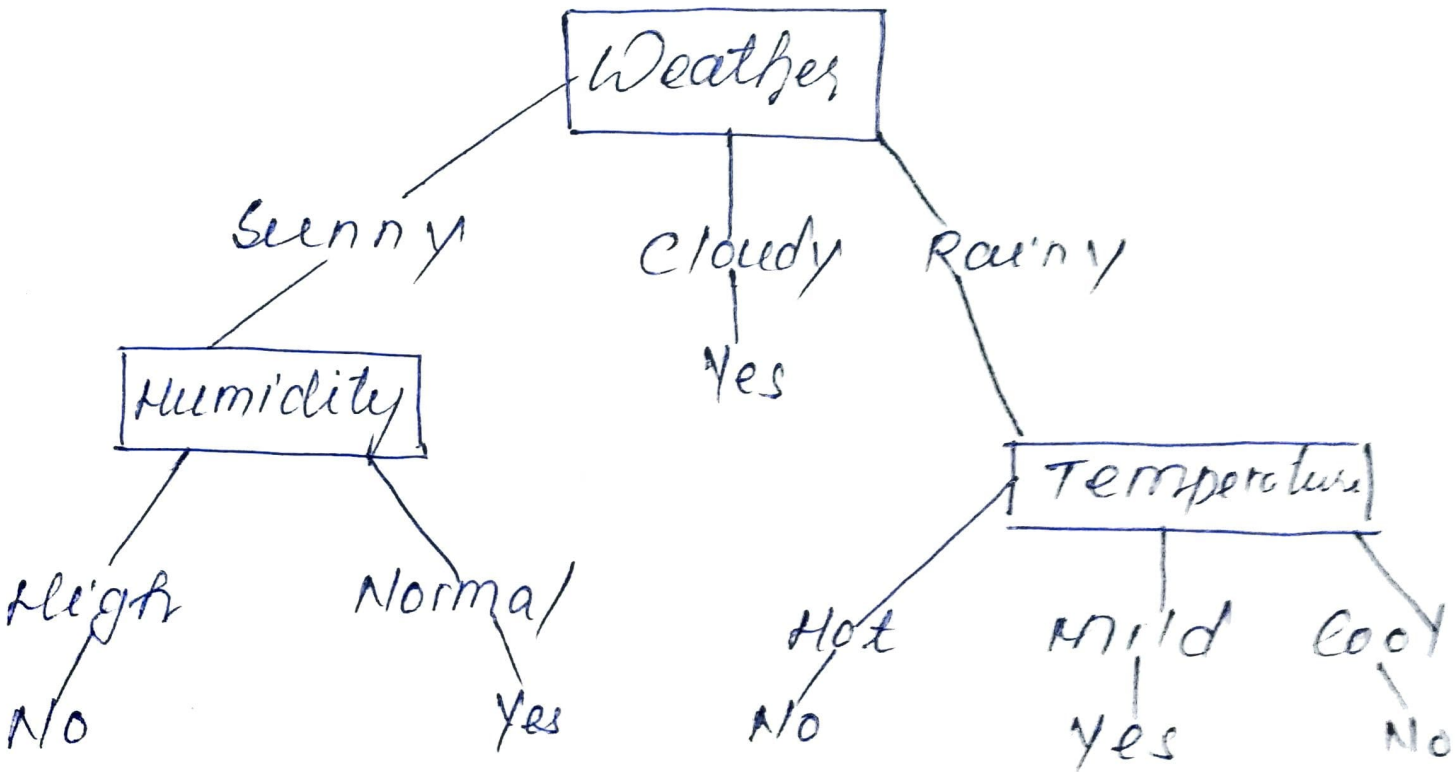

Fig: A decision tree for the condition of playing Badminton.