Table 6.14: Discretized Instances

S.No.	CGPA Continuous	CGPA Discretized	Job Offer
1.	9.5	>7.9	Yes
2.	8.2	>7.9	Yes
3.	9.1	>7.9	No
4.	6.8	≤7.9	No
5.	8.5	>7.9	Yes
6.	9.5	>7.9	Yes
7.	7.9	≤7.9	No
8.	9.1	>7.9	Yes
9.	8.8	>7.9	Yes
10.	8.8	>7.9	Yes

6.2.3 Classification and Regression Trees Construction

The Classification and Regression Trees (CART) algorithm is a multivariate decision tree learning used for classifying both categorical and continuous-valued target variables. CART algorithm is an example of multivariate decision trees that gives oblique splits. It solves both classification and regression problems. If the target feature is categorical, it constructs a classification tree and if the target feature is continuous, it constructs a regression tree. CART uses GINI Index to construct a decision tree. GINI Index is defined as the number of data instances for a class or it is the proportion of instances. It constructs the tree as a binary tree by recursively splitting a node into two nodes. Therefore, even if an attribute has more than two possible values, GINI Index is calculated for all subsets of the attributes and the subset which has maximum value is selected as the best split subset. For example, if an attribute A has three distinct values say $\{a_1, a_2, a_3\}$, the possible subsets are $\{\ \}$, $\{a_1\}$, $\{a_2\}$, $\{a_3\}$, $\{a_1, a_2\}$, $\{a_1, a_3\}$, $\{a_2, a_3\}$, and $\{a_1, a_2, a_3\}$. So, if an attribute has 3 distinct values, the number of possible subsets is 2^3, which means 8. Excluding the empty set $\{\ \}$ and the full set $\{a_1, a_2, a_3\}$, we have 6 subsets. With 6 subsets, we can form three possible combinations such as:

$\{a_1\}$ with $\{a_2, a_3\}$

$\{a_2\}$ with $\{a_1, a_3\}$

$\{a_3\}$ with $\{a_1, a_2\}$

Hence, in this CART algorithm, we need to compute the best splitting attribute and the best split subset i in the chosen attribute.

Higher the GINI value, higher is the homogeneity of the data instances.

Gini_Index(T) is computed as given in Eq. (6.13).

$$\text{Gini\_Index}(T) = 1 - \sum_{i=1}^{m} P_i^2 \qquad (6.13)$$

where,

P_i be the probability that a data instance or a tuple 'd' belongs to class C_i. It is computed as:

P_i = |No. of data instances belonging to class i|/|Total no of data instances in the training dataset T|

GINI Index assumes a binary split on each attribute, therefore, every attribute is considered as a binary attribute which splits the data instances into two subsets S_1 and S_2.

Gini_Index(T, A) is computed as given in Eq. (6.14).

$$\text{Gini\_Index}(T, A) = \frac{|S_1|}{|T|}\text{Gini}(S_1) + \frac{|S_2|}{|T|}\text{Gini}(S_2) \tag{6.14}$$

The splitting subset with minimum Gini_Index is chosen as the best splitting subset for an attribute. The best splitting attribute is chosen by the minimum Gini_Index which is otherwise maximum ΔGini because it reduces the impurity.

ΔGini is computed as given in Eq. (6.15):

$$\Delta\text{Gini}(A) = \text{Gini}(T) - \text{Gini}(T, A) \tag{6.15}$$

Algorithm 6.4: Procedure to Construct a Decision Tree using CART

1. Compute Gini_Index Eq. (6.13) for the whole training dataset based on the target attribute.
2. Compute Gini_Index for each of the attribute Eq. (6.14) and for the subsets of each attribute in the training dataset.
3. Choose the best splitting subset which has minimum Gini_Index for an attribute.
4. Compute ΔGini Eq. (6.15) for the best splitting subset of that attribute.
5. Choose the best splitting attribute that has maximum ΔGini.
6. The best split attribute with the best split subset is placed as the root node.
7. The root node is branched into two subtrees with each subtree an outcome of the test condition of the root node attribute. Accordingly, the training dataset is also split into two subsets.
8. Recursively apply the same operation for the subset of the training set with the remaining attributes until a leaf node is derived or no more training instances are available in the subset.

Example 6.5: Choose the same training dataset shown in Table 6.3 and construct a decision tree using CART algorithm.

Solution:

Step 1: Calculate the Gini_Index for the dataset shown in Table 6.3, which consists of 10 data instances. The target attribute 'Job Offer' has 7 instances as Yes and 3 instances as No.

$$\begin{aligned}\text{Gini\_Index}(T) &= 1 - \left(\frac{7}{10}\right)^2 - \left(\frac{3}{10}\right)^2 \\ &= 1 - 0.49 - 0.09 \\ &= 1 - 0.58 \\ \text{Gini\_Index}(T) &= 0.42\end{aligned}$$

Step 2: Compute Gini_Index for each of the attribute and each of the subset in the attribute.

CGPA has 3 categories, so there are 6 subsets and hence 3 combinations of subsets (as shown in Table 6.15).

Table 6.15: Categories of CGPA

CGPA	Job Offer = Yes	Job Offer = No
≥9	3	1
≥8	4	0
<8	0	2

$$\text{Gini\_Index}(T, \text{CGPA} \in \{\geq 9, \geq 8\}) = 1 - (7/8)^2 - (1/8)^2$$
$$= 1 - 0.7806$$
$$= 0.2194$$

$$\text{Gini\_Index}(T, \text{CGPA} \in \{<8\}) = 1 - (0/2)^2 - (2/2)^2$$
$$= 1 - 1$$
$$= 0$$

$$\text{Gini\_Index}(T, \text{CGPA} \in \{(\geq 9, \geq 8), <8\} = (8/10) \times 0.2194 + (2/10) \times 0$$
$$= 0.17552$$

$$\text{Gini\_Index}(T, \text{CGPA} \in \{\geq 9, <8\}) = 1 - (3/6)^2 - (3/6)^2$$
$$= 1 - 0.5 = 0.5$$

$$\text{Gini\_Index}(T, \text{CGPA} \in \{\geq 8\}) = 1 - (4/4)^2 - (0/4)^2$$
$$= 1 - 1 = 0$$

$$\text{Gini\_Index}(T, \text{CGPA} \in \{(\geq 9, <8), \geq 8\}) = (6/10) \times 0.5 + (4/10) \times 0$$
$$= 0.3$$

$$\text{Gini\_Index}(T, \text{CGPA} \in \{\geq 8, <8\}) = 1 - (4/6)^2 - (2/6)^2$$
$$= 1 - 0.555$$
$$= 0.445$$

$$\text{Gini\_Index}(T, \text{CGPA} \in \{\geq 9\}) = 1 - (3/4)^2 - (1/4)^2$$
$$= 1 - 0.625$$
$$= 0.375$$

$$\text{Gini\_Index}(T, \text{CGPA} \in \{(\geq 8, <8), \geq 9\}) = (6/10) \times 0.445 + (4/10) \times 0.375$$
$$= 0.417$$

Table 6.16 shows the Gini_Index for 3 subsets of CGPA.

Table 6.16: Gini_Index of CGPA

Subsets		Gini_Index
(≥9, ≥8)	<8	0.1755
(≥9, <8)	≥8	0.3
(≥8, <8)	≥9	0.417

Step 3: Choose the best splitting subset which has minimum Gini_Index for an attribute.

The subset CGPA ∈ {(≥9 ≥8), <8} has the lowest Gini_Index value as 0.1755 is chosen as the best splitting subset.

Step 4: Compute ΔGini or the best splitting subset of that attribute.

$$\begin{aligned}\Delta\text{Gini(CGPA)} &= \text{Gini}(T) - \text{Gini}(T, \text{CGPA})\\ &= 0.42 - 0.1755\\ &= 0.2445\end{aligned}$$

Repeat the same process for the remaining attributes in the dataset such as for Interactiveness shown in Table 6.17, Practical Knowledge in Table 6.18, and Communication Skills in Table 6.20.

Table 6.17: Categories for Interactiveness

Interactiveness	Job Offer = Yes	Job Offer = No
Yes	5	1
No	2	2

$$\begin{aligned}\text{Gini\_Index}(T, \text{Interactiveness} \in \{\text{Yes}\}) &= 1 - \left(\frac{5}{6}\right)^2 - \left(\frac{1}{6}\right)^2\\ &= 1 - 0.72\\ &= 0.28\end{aligned}$$

$$\begin{aligned}\text{Gini\_Index}(T, \text{Interactiveness} \in \{\text{No}\}) &= 1 - \left(\frac{2}{4}\right)^2 - \left(\frac{2}{4}\right)^2\\ &= 1 - 0.5\\ &= 0.5\end{aligned}$$

$$\begin{aligned}\text{Gini\_Index}(T, \text{Interactiveness} \in \{\text{Yes, No}\}) &= \frac{6}{10}(0.28) + \frac{4}{10}(0.5)\\ &= 0.168 + 0.2\\ &= 0.368\end{aligned}$$

$$\begin{aligned}\Delta\text{Gini(Interactiveness)} &= \text{Gini}(T) - \text{Gini}(T, \text{Interactiveness})\\ &= 0.42 - 0.368\\ &= 0.052\end{aligned}$$

Table 6.18: Categories for Practical Knowledge

Practical Knowledge	Job Offer = Yes	Job Offer = No
Very Good	2	0
Good	4	1
Average	1	2

$$\begin{aligned}\text{Gini\_Index}(T, \text{Practical Knowledge} \in \{\text{Very Good, Good}\} &= \left(\frac{6}{7}\right)^2 - \left(\frac{1}{7}\right)^2\\ &= 1 - 0.7544\\ &= 0.2456\end{aligned}$$

$$\begin{aligned}\text{Gini\_Index}(T, \text{Practical Knowledge} \in \{\text{Average}\}) &= 1 - \left(\frac{1}{3}\right)^2 - \left(\frac{2}{3}\right)^2\\ &= 1 - 0.555 = 0.445\end{aligned}$$

Gini_Index(*T*, Practical Knowledge ∈ {Very Good, Good}, Average)

$$= \left(\frac{7}{10}\right)^2 \times 0.2456 + \left(\frac{3}{10}\right) \times 0.445$$

$$= 0.3054$$

Gini_Index(*T*, Practical Knowledge ∈ {Very Good, Average}) $= 1 - \left(\frac{3}{5}\right)^2 - \left(\frac{2}{5}\right)^2$

$$= 1 - 0.52$$

$$= 0.48$$

Gini_Index(*T*, Practical Knowledge ∈ {Good}) $= 1 - \left(\frac{4}{5}\right)^2 - \left(\frac{1}{5}\right)^2$

$$= 1 - 0.68$$

$$= 0.32$$

Gini_Index(*T*, Practical Knowledge ∈ {Very Good, Average}, Good) $= \left(\frac{5}{10}\right) \times 0.48 + \left(\frac{5}{10}\right) \times 0.32$

$$= 0.40$$

Gini_Index(*T*, Practical Knowledge ∈ {~~Very~~ Good, Average}) $= 1 - \left(\frac{5}{8}\right)^2 - \left(\frac{3}{8}\right)^2$

$$= 1 - 0.5312 = 0.4688$$

Gini_Index(*T*, Practical Knowledge ∈ {Very Good}) $= 1 - \left(\frac{2}{2}\right)^2 - \left(\frac{0}{2}\right)^2$

$$= 1 - 1 = 0$$

Gini_Index(*T*, Practical Knowledge ∈ {Good, Average}, Very Good) $= \left(\frac{8}{10}\right) \times 0.4688 + \left(\frac{2}{10}\right) \times 0$

$$= 0.3750$$

Table 6.19 shows the Gini_Index for various subsets of Practical Knowledge.

Table 6.19: Gini_Index for Practical Knowledge

Subsets		Gini_Index
(Very Good, Good)	Average	0.3054
(Very Good, Average)	Good	0.40
(Good, Average)	Very Good	0.3750

ΔGini(Practical Knowledge) = Gini(*T*) – Gini(*T*, Practical Knowledge)

$$= 0.42 - 0.3054 = 0.1146$$

Table 6.20: Categories for Communication Skills

Communication Skills	Job Offer = Yes	Job Offer = No
Good	4	1
Moderate	3	0
Poor	0	2

$$\text{Gini\_Index}(T, \text{Communication Skills} \in \{\text{Good, Moderate}\}) = 1 - \left(\frac{7}{8}\right)^2 - \left(\frac{1}{8}\right)^2$$
$$= 1 - 0.7806$$
$$= 0.2194$$

$$\text{Gini\_Index}(T, \text{Communication Skills} \in \{\text{Poor}\}) = 1 - \left(\frac{2}{2}\right)^2 - \left(\frac{0}{2}\right)^2$$
$$= 1 - 1 = 0$$

$$\text{Gini\_Index}(T, \text{Communication Skills} \in \{\text{Good, Moderate}\}, \text{Poor}) = \left(\frac{8}{10}\right) \times 0.2194 + \left(\frac{2}{10}\right) \times 0$$
$$= 0.1755$$

$$\text{Gini\_Index}(T, \text{Communication Skills} \in \{\text{Good, Poor}\}) = 1 - \left(\frac{4}{7}\right)^2 - \left(\frac{3}{7}\right)^2$$
$$= 1 - 0.5101$$
$$= 0.4899$$

$$\text{Gini\_Index}(T, \text{Communication Skills} \in \{\text{Moderate}\}) = 1 - \left(\frac{3}{3}\right)^2 - \left(\frac{0}{3}\right)^2$$
$$= 1 - 1 = 0$$

$$\text{Gini\_Index}(T, \text{Communication Skills} \in \{\text{Good, Poor}\}, \text{Moderate}) = \left(\frac{7}{10}\right) \times 0.4899 + \left(\frac{3}{10}\right) \times 0$$
$$= 0.3429$$

$$\text{Gini\_Index}(T, \text{Communication Skills} \in \{\text{Moderate, Poor}\}) = 1 - \left(\frac{3}{5}\right)^2 - \left(\frac{2}{5}\right)^2$$
$$= 1 - 0.52$$
$$= 0.48$$

$$\text{Gini\_Index}(T, \text{Communication Skills} \in \{\text{Good}\}) = 1 - \left(\frac{4}{5}\right)^2 - \left(\frac{1}{5}\right)^2$$
$$= 1 - 0.68$$
$$= 0.32$$

$$\text{Gini\_Index}(T, \text{Communication Skills} \in \{\text{Moderate, Poor}\}, \text{Good}) = \left(\frac{5}{10}\right)^2 \times 048 + \left(\frac{5}{10}\right)^2 \times 0.32$$
$$= 0.40$$

Table 6.21 shows the Gini_Index for various subsets of Communication Skills.

Table 6.21: Gini-Index for Subsets of Communication Skills

Subsets		Gini_Index
(Good, Moderate)	Poor	0.1755
(Good, Poor)	Moderate	0.3429
(Moderate, Poor)	Good	0.40

$$\Delta\text{Gini(Communication Skills)} = \text{Gini}(T) - \text{Gini}(T, \text{Communication Skills})$$
$$= 0.42 - 0.1755$$
$$= 0.2445$$

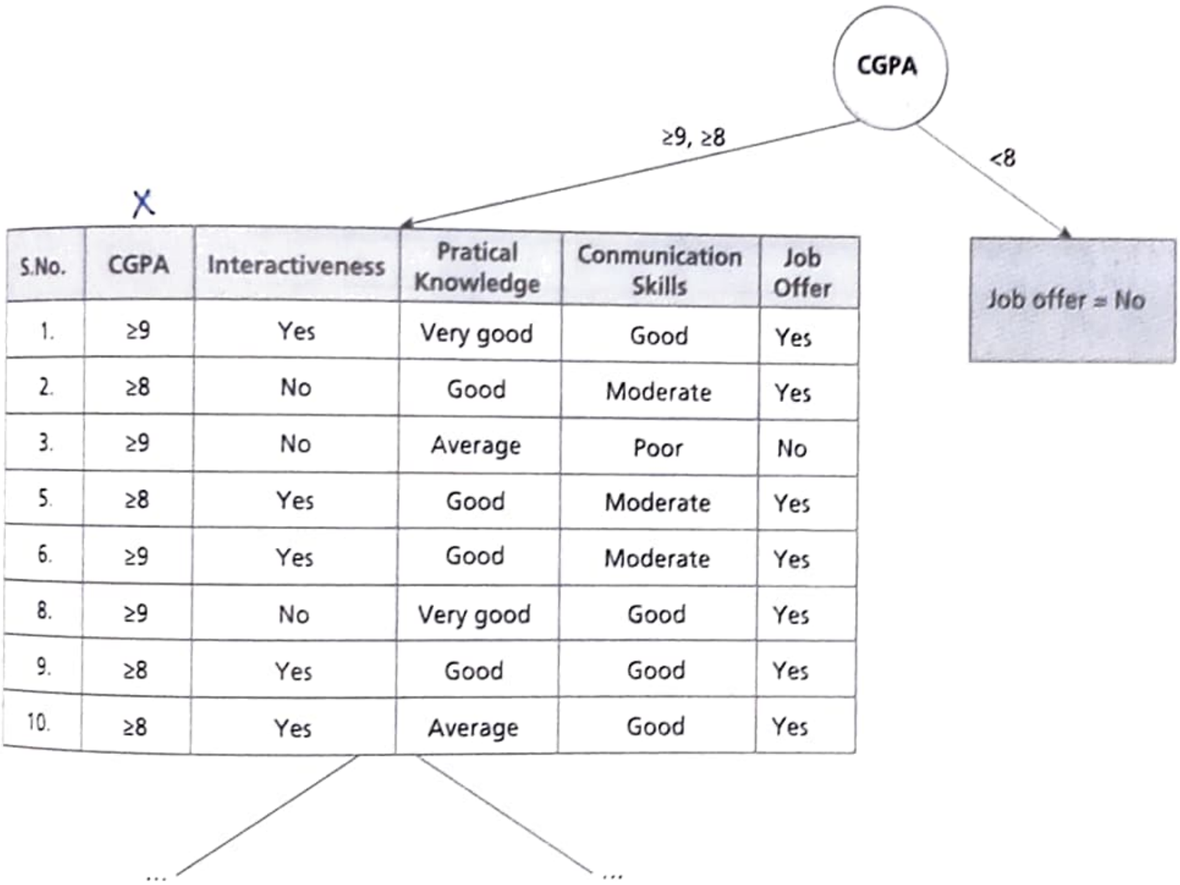
Table 6.22 shows the Gini_Index and ΔGini values calculated for all the attributes.

Table 6.22: Gini_Index and ΔGini for all Attributes

Attribute	Gini_Index	ΔGini
CGPA	0.1755	0.2445
Interactiveness	0.368	0.052
Practical knowledge	0.3054	0.1146
Communication Skills	0.1755	0.2445

Step 5: Choose the best splitting attribute that has maximum ΔGini.

CGPA and Communication Skills have the highest ΔGini value. We can choose CGPA as the root node and split the datasets into two subsets shown in Figure 6.7 since the tree constructed by CART is a binary tree.



S.No.	CGPA	Interactiveness	Pratical Knowledge	Conmunication Skills	Job Offer
1.	≥9	Yes	Very good	Good	Yes
2.	≥8	No	Good	Moderate	Yes
3.	≥9	No	Average	Poor	No
5.	≥8	Yes	Good	Moderate	Yes
6.	≥9	Yes	Good	Moderate	Yes
8.	≥9	No	Very good	Good	Yes
9.	≥8	Yes	Good	Good	Yes
10.	≥8	Yes	Average	Good	Yes

Figure 6.7: Decision Tree after Iteration 1

Iteration 2:

In the second iteration, the dataset has 8 data instances as shown in Table 6.23. Repeat the same process to find the best splitting attribute and the splitting subset for that attribute.

Table 6.23: Subset of the Training Dataset after Iteration 1

S.No.	CGPA	Interactiveness	Practical Knowledge	Communication Skills	Job Offer
1.	≥9	Yes	Very good	Good	Yes
2.	≥8	No	Good	Moderate	Yes
3.	≥9	No	Average	Poor	No
5.	≥8	Yes	Good	Moderate	Yes
6.	≥9	Yes	Good	Moderate	Yes
8.	≥9	No	Very good	Good	Yes
9.	≥8	Yes	Good	Good	Yes
10.	≥8	Yes	Average	Good	Yes

$$\text{Gini\_Index}(T) = 1 - \left(\frac{7}{8}\right)^2 - \left(\frac{1}{8}\right)^2$$
$$= 1 - 0.766 - 0.0156$$
$$= 1 - 0.58$$
$$\text{Gini\_Index}(T) = 0.2184$$

Tables 6.24, 6.25, and 6.27 show the categories for attributes Interactiveness, Practical Knowledge, and Communication Skills, respectively.

Table 6.24: Categories for Interactiveness

Interactiveness	Job Offer = Yes	Job Offer = No
Yes	5	0
No	2	1

$$\text{Gini\_Index}(T, \text{Interactiveness} \in \{\text{Yes}\}) = 1 - \left(\frac{5}{5}\right)^2 - \left(\frac{0}{5}\right)^2$$
$$= 1 - 1 = 0$$
$$\text{Gini\_Index}(T, \text{Interactiveness} \in \{\text{No}\}) = 1 - \left(\frac{2}{3}\right)^2 - \left(\frac{1}{3}\right)^2$$
$$= 1 - 0.44 - 0.111 = 0.449$$
$$\text{Gini\_Index}(T, \text{Interactiveness} \in \{\text{Yes, No}\}) = \left(\frac{7}{8}\right) \times 0 + \left(\frac{1}{8}\right) \times 0.449$$
$$= 0.056$$
$$\Delta\text{Gini(Interactiveness)} = \text{Gini}(T) - \text{Gini}(T, \text{Interactiveness})$$
$$= 0.2184 - 0.056 = 0.1624$$

Table 6.25: Categories for Practical Knowledge

Practical Knowledge	Job Offer = Yes	Job Offer = No
Very Good	2	0
Good	4	0
Average	1	1

$$\text{Gini\_Index}(T, \text{Practical Knowledge} \in \{\text{Very Good, Good}\}) = 1 - \left(\frac{6}{6}\right)^2 - \left(\frac{0}{6}\right)^2$$
$$= 1 - 1 = 0$$

$$\text{Gini\_Index}(T, \text{Practical Knowledge} \in \{\text{Average}\}) = 1 - \left(\frac{1}{2}\right)^2 - \left(\frac{1}{2}\right)^2$$
$$= 1 - 0.25 - 0.25$$
$$= 0.5$$

$$\text{Gini\_Index}(T, \text{Practical Knowledge} \in \{\text{Very Good, Good}\}, \text{Average}) = \left(\frac{6}{8}\right)^2 \times 0 + \left(\frac{2}{8}\right) \times 0.5$$
$$= 0.125$$

$$\text{Gini\_Index}(T, \text{Practical Knowledge} \in \{\text{Very Good, Average}\}) = 1 - \left(\frac{3}{4}\right)^2 - \left(\frac{1}{4}\right)^2$$
$$= 1 - 0.5625 - 0.0625$$
$$= 0.375$$

$$\text{Gini\_Index}(T, \text{Practical Knowledge} \in \{\text{Good}\}) = 1 - \left(\frac{4}{4}\right)^2 - \left(\frac{0}{4}\right)^2$$
$$= 1 - 1 = 0$$

$$\text{Gini\_Index}(T, \text{Practical Knowledge} \in \{\text{Very Good, Average}\}, \text{Good}) = \left(\frac{4}{8}\right) \times 0.375 + \left(\frac{4}{8}\right) \times 0$$
$$= 0.1875$$

$$\text{Gini\_Index}(T, \text{Practical Knowledge} \in \{\text{Good, Average}\}) = 1 - \left(\frac{5}{6}\right)^2 - \left(\frac{1}{6}\right)^2$$
$$= 1 - 0.694 - 0.028$$
$$= 0.278$$

$$\text{Gini\_Index}(T, \text{Practical Knowledge} \in \{\text{Very Good}\}) = 1 - \left(\frac{2}{2}\right)^2 - \left(\frac{0}{2}\right)^2$$
$$= 1 - 1 = 0$$

$$\text{Gini\_Index}(T, \text{Practical Knowledge} \in \{\text{Good, Average}\}, \text{Very Good}) = \left(\frac{6}{8}\right)^2 \times 0.278 + \left(\frac{2}{8}\right)^2 \times 0$$
$$= 0.2085$$

Table 6.26 shows the Gini_Index values for various subsets of Practical Knowledge.

Table 6.26: Gini_Index for Subsets of Practical Knowledge

Subsets		Gini_Index
{Very Good, Good}	Average	0.125
{Very Good, Average}	Good	0.1875
{Good, Average}	Very Good	0.2085

$$\Delta Gini(Practical\ Knowledge) = Gini(T) - Gini(T, Practical\ Knowledge)$$
$$= 0.2184 - 0.125$$
$$= 0.0934$$

Table 6.27: Categories for Communication Skills

Communication Skills	Job Offer = Yes	Job Offer = No
Good	4	0
Moderate	3	0
Poor	0	1

$$Gini\_Index(T, Communication\ Skills \in \{Good, Moderate\}) = 1 - \left(\frac{7}{7}\right)^2 - \left(\frac{0}{7}\right)^2$$
$$= 1 - 1 = 0$$

$$Gini\_Index(T, Communication\ Skills \in \{Poor\}) = 1 - \left(\frac{0}{1}\right)^2 - \left(\frac{1}{1}\right)^2$$
$$= 1 - 1 = 0$$

$$Gini\_Index(T, Communication\ Skills \in \{Good, Moderate\}, Poor) = \left(\frac{7}{8}\right)^2 \times 0 + \left(\frac{1}{8}\right) \times 0$$
$$= 0$$

$$Gini\_Index(T, Communication\ Skills \in \{Good, Poor\}) = 1 - \left(\frac{4}{5}\right)^2 - \left(\frac{1}{5}\right)^2$$
$$= 1 - 0.64 - 0.04$$
$$= 0.32$$

$$Gini\_Index(T, Communication\ Skills \in \{Moderate\}) = 1 - \left(\frac{3}{3}\right)^2 - \left(\frac{0}{3}\right)^2$$
$$= 1 - 1 = 0$$

$$Gini\_Index(T, Communication\ Skills \in \{Good, Poor\}, Moderate) = \left(\frac{5}{8}\right) \times 0.32 + \left(\frac{3}{8}\right) \times 0$$
$$= 0.2$$

$$Gini\_Index(T, Communication\ Skills \in \{Moderate, Poor\}) = 1 - \left(\frac{3}{4}\right)^2 - \left(\frac{1}{4}\right)^2$$
$$= 1 - 0.5625 - 0.0625$$
$$= 0.375$$

$$Gini\_Index(T, Communication\ Skills \in \{Good\}) = 1 - \left(\frac{4}{4}\right)^2 - \left(\frac{0}{4}\right)^2$$
$$= 1 - 1 = 0$$

$$Gini\_Index(T, Communication\ Skills \in \{Moderate, Poor\}, Good) = \left(\frac{4}{8}\right)^2 \times 0.375 + \left(\frac{4}{8}\right)^2 \times 0$$
$$= 0.1875$$

Table 6.28 shows the Gini_Index for subsets of Communication Skills.

Table 6.28: Gini_Index for Subsets of Communication Skills

Subsets		Gini_Index
(Good, Moderate)	Poor	0
(Good, Poor)	Moderate	0.2
(Moderate, Poor)	Good	0.1875

$$\Delta Gini(\text{Communication Skills}) = Gini(T) - Gini(T, \text{Communication Skills})$$
$$= 0.2184 - 0 = 0.2184$$

Table 6.29 shows the Gini_Index and ΔGini values for all attributes.

Table 6.29: Gini_Index and ΔGini Values for All Attributes

Attribute	Gini_Index	ΔGini
Interactiveness	0.056	0.1624
Practical knowledge	0.125	0.0934
Communication Skills	0	0.2184

Communication Skills has the highest ΔGini value. The tree is further branched based on the attribute 'Communication Skills'. Here, we see all branches end up in a leaf node and the process of construction is completed. The final tree is shown in Figure 6.8.

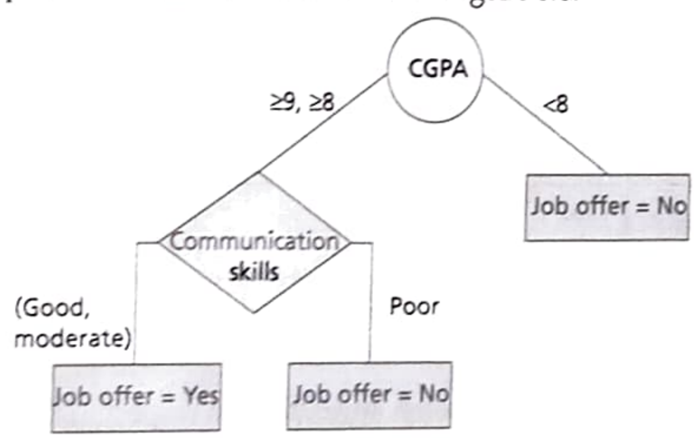


Figure 6.8: Final Tree

6.2.4 Regression Trees

Regression trees are a variant of decision trees where the target feature is a continuous valued variable. These trees can be constructed using an algorithm called reduction in variance which uses standard deviation to choose the best splitting attribute.

Algorithm 6.5: Procedure for Constructing Regression Trees

1. Compute standard deviation for each attribute with respect to target attribute.
2. Compute standard deviation for the number of data instances of each distinct value of an attribute.
3. Compute weighted standard deviation for each attribute.

(Continued)