Decision Tree Worked Example

Contents

1 Problem Statement

We consider a binary classification task where the goal is to predict whether a student *passes* an exam:

$$\text{Pass} \in \{\text{Yes}, \text{No}\}.$$

The prediction is based on three categorical attributes:

- StudyHours $\in \{\text{Low}, \text{Medium}, \text{High}\}$,
- Attendance $\in \{\text{Poor}, \text{Good}\}$,
- Sleep $\in \{\text{Low}, \text{Normal}\}$.

The training dataset is:

ID	StudyHours	Attendance	Sleep	Pass
1	Low	Poor	Low	No
2	Low	Good	Low	No
3	Medium	Good	Low	Yes
4	Medium	Good	Normal	Yes
5	Medium	Poor	Normal	No
6	High	Good	Low	Yes
7	High	Poor	Normal	Yes
8	High	Good	Normal	Yes

We will:

1. Compute the entropy $H(\text{Pass})$ of the target.
2. Compute the information gain for each attribute (StudyHours, Attendance, Sleep).
3. Choose the root attribute and build the decision tree (at least two levels).
4. Use the resulting tree to classify a new example:

$$(\text{StudyHours} = \text{Medium},\ \text{Attendance} = \text{Poor},\ \text{Sleep} = \text{Low}).$$

2 Entropy of the Target Variable $H(\text{Pass})$

We first count how many examples belong to each class:

- Pass = Yes: IDs 3, 4, 6, 7, 8 $\Rightarrow$ 5 examples.
- Pass = No: IDs 1, 2, 5 $\Rightarrow$ 3 examples.

Total number of examples:

$$N = 8.$$

Class probabilities:

$$P(\text{Yes}) = \frac{5}{8}, \qquad P(\text{No}) = \frac{3}{8}.$$

The entropy of a binary variable Y with probabilities p and $1-p$ is

$$H(Y) = -\left[p \log_2 p + (1-p)\log_2(1-p)\right].$$

For our case:

$$H(\text{Pass}) = -\left[\frac{5}{8}\log_2\left(\frac{5}{8}\right) + \frac{3}{8}\log_2\left(\frac{3}{8}\right)\right].$$

We can write this explicitly as:

$$H(\text{Pass}) = -\frac{5}{8}\log_2\left(\frac{5}{8}\right) - \frac{3}{8}\log_2\left(\frac{3}{8}\right) \approx 0.9544 \text{ bits}.$$

For later calculations, we keep:

$$H(\text{Pass}) \approx 0.9544.$$

3 Information Gain for Each Attribute

The information gain for an attribute X with respect to the target Y is defined as:

$$IG(X) = H(Y) - H(Y \mid X),$$

where $H(Y \mid X)$ is the conditional entropy:

$$H(Y \mid X) = \sum_x P(X = x)\, H(Y \mid X = x).$$

In our setting, $Y = \text{Pass}$ and X is one of the attributes: StudyHours, Attendance, or Sleep.

3.1 Information Gain for StudyHours

The attribute StudyHours takes three values:

$$\{\text{Low}, \text{Medium}, \text{High}\}.$$

We partition the data by StudyHours.

Case 1: StudyHours = Low

Rows with StudyHours = Low: IDs 1, 2.

ID	StudyHours	Attendance	Sleep	Pass
1	Low	Poor	Low	No
2	Low	Good	Low	No

Counts:

$$\#\text{Yes} = 0, \quad \#\text{No} = 2.$$

Probabilities:

$$P(\text{Yes} \mid \text{Low}) = 0, \quad P(\text{No} \mid \text{Low}) = 1.$$

Entropy:

$$H(\text{Pass} \mid \text{StudyHours} = \text{Low}) = 0$$

because the subset is pure (all "No").

Case 2: StudyHours = Medium

Rows: IDs 3, 4, 5.

ID	StudyHours	Attendance	Sleep	Pass
3	Medium	Good	Low	Yes
4	Medium	Good	Normal	Yes
5	Medium	Poor	Normal	No

Counts:

$$\#\text{Yes} = 2, \quad \#\text{No} = 1, \quad \text{total} = 3.$$

Probabilities:

$$P(\text{Yes} \mid \text{Medium}) = \frac{2}{3}, \qquad P(\text{No} \mid \text{Medium}) = \frac{1}{3}.$$

Entropy:

$$\begin{aligned} H(\text{Pass} \mid \text{StudyHours} = \text{Medium}) &= -\left[\frac{2}{3}\log_2\left(\frac{2}{3}\right) + \frac{1}{3}\log_2\left(\frac{1}{3}\right)\right] \\ &\approx 0.9183. \end{aligned}$$

Case 3: StudyHours = High

Rows: IDs 6, 7, 8.

ID	StudyHours	Attendance	Sleep	Pass
6	High	Good	Low	Yes
7	High	Poor	Normal	Yes
8	High	Good	Normal	Yes

Counts:

$$\#\text{Yes} = 3, \quad \#\text{No} = 0,$$

so this subset is also pure and

$$H(\text{Pass} \mid \text{StudyHours} = \text{High}) = 0.$$

Weighted Conditional Entropy for StudyHours

Subset sizes:

$$|\text{Low}| = 2, \quad |\text{Medium}| = 3, \quad |\text{High}| = 3, \quad N = 8.$$

Thus:

$$P(\text{Low}) = \frac{2}{8}, \quad P(\text{Medium}) = \frac{3}{8}, \quad P(\text{High}) = \frac{3}{8}.$$

Conditional entropy:

$$\begin{aligned}
H(\text{Pass} \mid \text{StudyHours}) &= P(\text{Low}) \cdot H(\text{Pass} \mid \text{Low}) + P(\text{Medium}) \cdot H(\text{Pass} \mid \text{Medium}) + P(\text{High}) \cdot H(\text{Pass} \mid \text{High}\\
&= \frac{2}{8} \cdot 0 + \frac{3}{8} \cdot 0.9183 + \frac{3}{8} \cdot 0 \\
&\approx \frac{3}{8} \cdot 0.9183 \\
&\approx 0.3444.
\end{aligned}$$

Information Gain for StudyHours

$$\begin{aligned}
IG(\text{StudyHours}) &= H(\text{Pass}) - H(\text{Pass} \mid \text{StudyHours}) \\
&\approx 0.9544 - 0.3444 \\
&\approx 0.6101.
\end{aligned}$$

3.2 Information Gain for Attendance

Attendance takes two values: $\{\text{Poor}, \text{Good}\}$.

Case 1: Attendance = Poor

Rows: IDs 1, 5, 7.

ID	StudyHours	Attendance	Sleep	Pass
1	Low	Poor	Low	No
5	Medium	Poor	Normal	No
7	High	Poor	Normal	Yes

Counts:

$$\#\text{Yes} = 1, \quad \#\text{No} = 2, \quad \text{total} = 3.$$

Probabilities:

$$P(\text{Yes} \mid \text{Poor}) = \frac{1}{3}, \qquad P(\text{No} \mid \text{Poor}) = \frac{2}{3}.$$

Entropy:

$$\begin{aligned} H(\text{Pass} \mid \text{Attendance} = \text{Poor}) &= -\left[\frac{1}{3}\log_2\left(\frac{1}{3}\right) + \frac{2}{3}\log_2\left(\frac{2}{3}\right)\right] \\ &\approx 0.9183. \end{aligned}$$

Case 2: Attendance = Good

Rows: IDs 2, 3, 4, 6, 8.

ID	StudyHours	Attendance	Sleep	Pass
2	Low	Good	Low	No
3	Medium	Good	Low	Yes
4	Medium	Good	Normal	Yes
6	High	Good	Low	Yes
8	High	Good	Normal	Yes

Counts:

$$\#\text{Yes} = 4, \quad \#\text{No} = 1, \quad \text{total} = 5.$$

Probabilities:

$$P(\text{Yes} \mid \text{Good}) = \frac{4}{5}, \qquad P(\text{No} \mid \text{Good}) = \frac{1}{5}.$$

Entropy:

$$\begin{aligned} H(\text{Pass} \mid \text{Attendance} = \text{Good}) &= -\left[\frac{4}{5}\log_2\left(\frac{4}{5}\right) + \frac{1}{5}\log_2\left(\frac{1}{5}\right)\right] \\ &\approx 0.7219. \end{aligned}$$

Weighted Conditional Entropy for Attendance

Subset sizes:

$$|\text{Poor}| = 3, \quad |\text{Good}| = 5, \quad N = 8.$$

Thus:

$$P(\text{Poor}) = \frac{3}{8}, \qquad P(\text{Good}) = \frac{5}{8}.$$

Conditional entropy:

$$\begin{aligned} H(\text{Pass} \mid \text{Attendance}) &= \frac{3}{8} \cdot 0.9183 + \frac{5}{8} \cdot 0.7219 \\ &\approx 0.3444 + 0.4512 \\ &\approx 0.7956. \end{aligned}$$

Information Gain for Attendance

$$\begin{aligned} IG(\text{Attendance}) &= H(\text{Pass}) - H(\text{Pass} \mid \text{Attendance}) \\ &\approx 0.9544 - 0.7956 \\ &\approx 0.1589. \end{aligned}$$

3.3 Information Gain for Sleep

Sleep takes two values: $\{\text{Low}, \text{Normal}\}$.

Case 1: Sleep = Low

Rows: IDs 1, 2, 3, 6.

ID	StudyHours	Attendance	Sleep	Pass
1	Low	Poor	Low	No
2	Low	Good	Low	No
3	Medium	Good	Low	Yes
6	High	Good	Low	Yes

Counts:

$$\#\text{Yes} = 2, \quad \#\text{No} = 2, \quad \text{total} = 4.$$

Probabilities:

$$P(\text{Yes} \mid \text{Low}) = \frac{1}{2}, \qquad P(\text{No} \mid \text{Low}) = \frac{1}{2}.$$

Entropy:

$$\begin{aligned} H(\text{Pass} \mid \text{Sleep} = \text{Low}) &= -\left[\frac{1}{2}\log_2\left(\frac{1}{2}\right) + \frac{1}{2}\log_2\left(\frac{1}{2}\right)\right] \\ &= 1. \end{aligned}$$

Case 2: Sleep = Normal

Rows: IDs 4, 5, 7, 8.

ID	StudyHours	Attendance	Sleep	Pass
4	Medium	Good	Normal	Yes
5	Medium	Poor	Normal	No
7	High	Poor	Normal	Yes
8	High	Good	Normal	Yes

Counts:

$$\#\text{Yes} = 3, \quad \#\text{No} = 1, \quad \text{total} = 4.$$

Probabilities:

$$P(\text{Yes} \mid \text{Normal}) = \frac{3}{4}, \qquad P(\text{No} \mid \text{Normal}) = \frac{1}{4}.$$

Entropy:

$$\begin{aligned} H(\text{Pass} \mid \text{Sleep} = \text{Normal}) &= -\left[\frac{3}{4}\log_2\left(\frac{3}{4}\right) + \frac{1}{4}\log_2\left(\frac{1}{4}\right)\right] \\ &\approx 0.8113. \end{aligned}$$

Weighted Conditional Entropy for Sleep

Subset sizes:

$$|\text{Low}| = 4, \quad |\text{Normal}| = 4, \quad N = 8.$$

Thus:

$$P(\text{Low}) = \frac{4}{8} = \frac{1}{2}, \qquad P(\text{Normal}) = \frac{4}{8} = \frac{1}{2}.$$

Conditional entropy:

$$\begin{aligned} H(\text{Pass} \mid \text{Sleep}) &= \frac{1}{2} \cdot 1 + \frac{1}{2} \cdot 0.8113 \\ &\approx 0.5 + 0.4056 \\ &\approx 0.9056. \end{aligned}$$

Information Gain for Sleep

$$\begin{aligned} IG(\text{Sleep}) &= H(\text{Pass}) - H(\text{Pass} \mid \text{Sleep}) \\ &\approx 0.9544 - 0.9056 \\ &\approx 0.0488. \end{aligned}$$

3.4 Summary of Information Gains

We summarize:

$$IG(\text{StudyHours}) \approx 0.6101,$$

$$IG(\text{Attendance}) \approx 0.1589,$$

$$IG(\text{Sleep}) \approx 0.0488.$$

The attribute with the largest information gain is:

$$\text{StudyHours},$$

so StudyHours is chosen as the **root** of the decision tree.

4 Choosing the Root and Building the Tree

4.1 Root Split on StudyHours

From the previous calculations, the top-level split is:

$$\text{Root attribute} = \text{StudyHours}.$$

We already know the class distributions in each branch:

- StudyHours = Low: all examples have Pass = No.
- StudyHours = High: all examples have Pass = Yes.
- StudyHours = Medium: mixed (2 Yes, 1 No).

Therefore, we can immediately assign leaves for two of the branches:

- StudyHours = Low $\Rightarrow$ leaf with prediction Pass = No.
- StudyHours = High $\Rightarrow$ leaf with prediction Pass = Yes.

The branch StudyHours = Medium is not pure, so we must split it further.

4.2 Subtree for StudyHours = Medium

We restrict attention to the subset where StudyHours = Medium:

ID	StudyHours	Attendance	Sleep	Pass
3	Medium	Good	Low	Yes
4	Medium	Good	Normal	Yes
5	Medium	Poor	Normal	No

Within this subset:

$$\#\text{Yes} = 2, \quad \#\text{No} = 1,$$

so the entropy is

$$H_{\text{parent}} = H(\text{Pass} \mid \text{StudyHours} = \text{Medium}) \approx 0.9183.$$

We compare splitting this subset on Attendance versus Sleep.

Split on Attendance (within StudyHours = Medium)

Possible values here: $\{\text{Good}, \text{Poor}\}$.

- Attendance = Good: IDs 3, 4 (both Pass = Yes) $\Rightarrow$ pure Yes, entropy 0.
- Attendance = Poor: ID 5 (Pass = No) $\Rightarrow$ pure No, entropy 0.

Weighted conditional entropy:

$$H(\text{Pass} \mid \text{Attendance}, \text{StudyHours} = \text{Medium}) = \frac{2}{3} \cdot 0 + \frac{1}{3} \cdot 0 = 0.$$

Information gain:

$$IG(\text{Attendance} \mid \text{StudyHours} = \text{Medium}) = H_{\text{parent}} - 0 \approx 0.9183.$$

Split on Sleep (within StudyHours = Medium)

Possible values here: $\{\text{Low}, \text{Normal}\}$.

- Sleep = Low: ID 3 (Pass = Yes) $\Rightarrow$ pure Yes, entropy 0.
- Sleep = Normal: IDs 4, 5 (Yes, No) $\Rightarrow$ one of each, entropy 1.

Weighted conditional entropy:

$$\begin{aligned} H(\text{Pass} \mid \text{Sleep}, \text{StudyHours} = \text{Medium}) &= \frac{1}{3} \cdot 0 + \frac{2}{3} \cdot 1 \\ &= \frac{2}{3} \approx 0.6667. \end{aligned}$$

Information gain:

$$IG(\text{Sleep} \mid \text{StudyHours} = \text{Medium}) \approx 0.9183 - 0.6667 \approx 0.2516.$$

Choice of Next Attribute

Since

$$IG(\text{Attendance} \mid \text{Medium}) \approx 0.9183 > IG(\text{Sleep} \mid \text{Medium}) \approx 0.2516,$$

we choose Attendance as the next split under StudyHours = Medium.

Under this split:

- StudyHours = Medium, Attendance = Good: IDs 3,4 $\Rightarrow$ all Pass = Yes.
- StudyHours = Medium, Attendance = Poor: ID 5 $\Rightarrow$ all Pass = No.

So these become leaf nodes.

4.3 Final Decision Tree (Text Description)

The tree can be described as follows:

- If StudyHours = Low, then predict Pass = No.
- If StudyHours = High, then predict Pass = Yes.
- If StudyHours = Medium, then:
  - If Attendance = Good, predict Pass = Yes.
  - If Attendance = Poor, predict Pass = No.

In pseudocode:

```
if StudyHours = Low:
    predict Pass = No
elif StudyHours = High:
    predict Pass = Yes
else:  # StudyHours = Medium
    if Attendance = Good:
        predict Pass = Yes
    else:  # Attendance = Poor
        predict Pass = No
```

Note that Sleep does not appear in the final tree; it did not provide as much information gain as the other attributes.

5 Classifying a New Example

We now classify the instance:

$$(\text{StudyHours} = \text{Medium},\ \text{Attendance} = \text{Poor},\ \text{Sleep} = \text{Low}).$$

We follow the decision tree from the root to a leaf:

1. **Root test:** Check StudyHours.
   - The instance has StudyHours = Medium, so we follow the branch:

     $$\text{StudyHours} = \text{Medium}.$$

2. **Second-level test (under StudyHours = Medium):** Check Attendance.

- The instance has Attendance = Poor, so we follow the branch:

$$\text{Attendance} = \text{Poor}.$$

3. **Leaf:** The leaf corresponding to

$$\text{StudyHours} = \text{Medium}, \quad \text{Attendance} = \text{Poor}$$

predicts:

$$\text{Pass} = \text{No}.$$

Therefore, the decision tree classifies the new instance as:

$$\boxed{\text{Pass} = \text{No}}.$$

Note that the attribute Sleep is *not* used in this path; once we know StudyHours and Attendance, the tree can already make a confident prediction.