1 Question-1

Prove that:

$$H(X,Y) = H(X) + H(Y|X)$$

Do not assume independence of X and Y.

Solution

$$H(X,Y) = -\sum_{x \in X}\sum_{y \in Y} p(x,y)log_2(p(x,y)) \tag{1}$$

Since variable independence cannot be assumed:

$$p(x,y) = p(x)p(y|x) \tag{2}$$

Substituting from Equation 2 in Equation 1:

$$H(X,Y) = -\sum_{x \in X}\sum_{y \in Y} p(x)p(y|x)log_2(p(x)p(y|x))$$

$$H(X,Y) = -\sum_{x \in X}\sum_{y \in Y} p(x)p(y|x)log_2(p(x)) - \sum_{x \in X}\sum_{y \in Y} p(x)p(y|x)log_2(p(y|x))$$

$$H(X,Y) = -\sum_{x \in X} p(x)log_2(p(x))\sum_{y \in Y} p(y|x) - \sum_{x \in X} p(x)\sum_{y \in Y} p(y|x)log_2(p(y|x)) \tag{3}$$

Now:

$$H(X) = -\sum_{x \in X} p(x)log_2(p(x))$$

$$H(Y|X) = -\sum_{x \in X} p(x)\sum_{y \in Y} p(y|x)log_2(p(y|x))$$

and:

$$\sum_{y \in Y} p(y|x) = 1$$

Substituting values in Equation 3:

◀ $$H(X,Y) = H(X) + H(Y|X)$$

2 Questtion-2

Discretize the following decision tables:

Cases	Attributes				Decision
	Length	Height	Width	Weight	Quality
1	4.7	1.8	1.7	1.7	high
2	4.5	1.4	1.8	0.9	high
3	4.7	1.8	1.9	1.3	high
4	4.5	1.8	1.7	1.3	medium
5	4.3	1.6	1.9	1.7	medium
6	4.3	1.4	1.7	0.9	low
7	4.5	1.6	1.9	0.9	very-low
8	4.5	1.4	1.8	1.3	very-low

Table 1 Dataset

Solution:

$$\{A\}^* = \{\{1\},\{2\},\{3\},\{4\},\{5\},\{6\},\{7\},\{8\}\}$$
$$\{d\}^* = \{\{1,2,3\},\{4,5\},\{6\},\{7,8\}\}$$
$$\{A\}^* \leq \{d\}^*$$

▶ Dataset is consistent!

Part-(a) - Dominant Attribute Approach

Trying Length:

$$Cases(5,6) = (medium, low)$$
$$Cases(2,4,7,8) = (high, medium, very-low, very-low)$$
$$Cases(1,3) = (high, high)$$
$$\triangleright H(Quality|Length) = -\frac{2}{8}(\frac{1}{2}log_2(\frac{1}{2}))2 - \frac{4}{8}(\frac{1}{4}log_2(\frac{1}{4})2 + \frac{2}{4}log_2(\frac{2}{4})) - \frac{2}{8}(0)$$
$$= 1$$

Trying Height:

$$Cases(2,6,8) = (high, low, very-low)$$
$$Cases(5,7) = (medium, very-low)$$
$$Cases(1,3,4) = (high, high, medium)$$
$$H(Quality|Height) = -\frac{3}{8}(\frac{1}{3}log_2(\frac{1}{3}))3 - \frac{2}{8}(-1) - \frac{3}{8}(\frac{2}{3}log_2(\frac{2}{3}) + \frac{1}{3}log_2(\frac{1}{3}))$$
$$= 1.189$$

Trying Width:

$$\begin{aligned} Cases(1,4,6)) &= (high, medium, low) \\ Cases(2,8) &= (high, very-low) \\ Cases(3,5,7) &= (high, medium, very-low) \\ H(Quality|Width) &= -\frac{3}{8}(\frac{1}{3}log_2(\frac{1}{3}))3 - \frac{2}{8}(-1) - \frac{3}{8}(\frac{1}{3}log_2(\frac{1}{3}))3 \\ &= 1.439 \end{aligned}$$

Trying Weight:

$$\begin{aligned} Cases(2,6,7) &= (high, low, very-low) \\ Cases(3,4,8) &= (high, medium, very-low) \\ Cases(1,5) &= (high, medium) \\ H(Quality|Weight) &= -\frac{3}{8}(\frac{1}{3}log_2(\frac{1}{3}))3 - \frac{3}{8}(\frac{1}{3}log_2(\frac{1}{3}))3 - \frac{2}{8}(-1) \\ &= 1.439 \end{aligned}$$

Dominant Attribute: Length
Cutpoints for Length: 4.4, 4.6

Trying $Length_{4.4}$**:**

$$\begin{aligned} Length_{4.3-4.4} &= Cases(5,6) = (medium, low) \\ Length_{4.4-4.7} &= Cases(1,2,3,4,7,8) = (high, high, high, medium, very-low, very-low) \\ H(Quality|Length_{4.4}) &= -\frac{2}{8}(-1) - \frac{6}{8}(\frac{3}{6}log_2(\frac{3}{6}) + \frac{1}{6}log_2(\frac{1}{6}) + \frac{2}{6}log_2(\frac{2}{6})) \\ &= 1.344 \end{aligned}$$

Trying $Length_{4.6}$**:**

$$\begin{aligned} Length_{4.3-4.6} &= Cases(2,4,5,6,7,8) = (high, medium, medium, low, very-low, very-low) \\ Length_{4.6-4.7} &= Cases(1,3) = (high, high) \\ H(Quality|Length_{4.6}) &= -\frac{6}{8}((\frac{1}{6}log_2(\frac{1}{6}))2 + (\frac{2}{6}log_2(\frac{2}{6}))2) - \frac{2}{8}(0) \\ &= 1.439 \end{aligned}$$

Cutpoint 4.4 is better for *Length*.

Cases	Attributes	Decision
	Length	Quality
5	4.3 - 4.4	medium
6	4.3 - 4.4	low
1	4.4 - 4.7	high
2	4.4 - 4.7	high
3	4.4 - 4.7	high
4	4.4 - 4.7	medium
7	4.4 - 4.7	very-low
8	4.4 - 4.7	very-low

Table 2 Discretized Dataset with $Length_{4.4}$

$$\{A^d\}^* = \{\{5,6\},\{1,2,3,4,7,8\}\}$$
$$\{d\}^* = \{\{1,2,3\},\{4,5\},\{6\},\{7,8\}\}$$
$$\{A\}^* \not\leq \{d\}^*$$

▷ Dataset is not consistent.

Cases	Attributes				Decision
	Length	Height	Width	Weight	Quality
5	4.3	1.6	1.9	1.7	medium
6	4.3	1.4	1.7	0.9	low

Table 3 Incosistent Cases after Discretization of $Length_{4.4}$

Dominant Attribute: Height
Cutpoints for Height: 1.5

Cases	Attributes		Decision
	Length	Height	Quality
5	4.3 - 4.4	1.5-1.9	medium
6	4.3 - 4.4	1.4-1.5	low
2	4.4 - 4.7	1.4-1.5	high
8	4.4 - 4.7	1.4-1.5	very-low
1	4.4 - 4.7	1.5-1.9	high
3	4.4 - 4.7	1.5-1.9	high
4	4.4 - 4.7	1.5-1.9	medium
7	4.4 - 4.7	1.5-1.9	very-low

Table 4 Discretized Dataset with $Length_{4.4}$ and $Height_{1.5}$

$$\{A^d\}^* = \{\{5\},\{6\},\{2,8\},\{1,3,4,7\}\}$$
$$\{d\}^* = \{\{1,2,3\},\{4,5\},\{6\},\{7,8\}\}$$
$$\{A\}^* \not\leq \{d\}^*$$

▷ Dataset is not consistent.

Cases	Attributes				Decision
	Length	Height	Width	Weight	Quality
2	4.5	1.4	1.8	0.9	high
8	4.5	1.4	1.8	1.3	very-low

Table 5 Inconsistent Cases after Discretization of *Length* and *Height*

Dominant Attribute: Weight
Cutpoints for Weight: 1.1

Cases	Attributes			Decision
	Length	Height	Weight	Quality
5	4.3 - 4.4	1.5-1.9	1.1-1.7	medium
6	4.3 - 4.4	1.4-1.5	0.9-1.1	low
2	4.4 - 4.7	1.4-1.5	0.9-1.1	high
8	4.4 - 4.7	1.4-1.5	1.1-1.7	very-low
1	4.4 - 4.7	1.5-1.9	1.1-1.7	high
3	4.4 - 4.7	1.5-1.9	1.1-1.7	high
4	4.4 - 4.7	1.5-1.9	1.1-1.7	medium
7	4.4 - 4.7	1.5-1.9	0.9-1.1	very-low

Table 6 Discretized Dataset with $Length_{4.4}$, $Height_{1.5}$ and $Weight_{1.1}$

$$\{A^d\}^* = \{\{5\},\{6\},\{2\},\{8\},\{1,3,4\},\{7\}\}$$
$$\{d\}^* = \{\{1,2,3\},\{4,5\},\{6\},\{7,8\}\}$$
$$\{A\}^* \not\leq \{d\}^*$$

▷ Dataset is not consistent.

Cases	Attributes				Decision
	Length	Height	Width	Weight	Quality
1	4.7	1.8	1.7	1.7	high
3	4.7	1.8	1.9	1.3	high
4	4.5	1.8	1.7	1.3	medium

Table 7 Inconsistent Cases after Discretization of *Length*, *Height* and *Weight*

Dominant Attribute: Length
Cutpoints for Length: 4.6

Cases	Attributes			Decision
	Length	Height	Weight	Quality
6	4.3 - 4.4	1.4-1.5	0.9-1.1	low
5	4.3 - 4.4	1.5-1.9	1.1-1.7	medium
2	4.4 - 4.6	1.4-1.5	0.9-1.1	high
8	4.4 - 4.6	1.4-1.5	1.1-1.7	very-low
4	4.4 - 4.6	1.5-1.9	1.1-1.7	medium
7	4.4 - 4.6	1.5-1.9	0.9-1.1	very-low
1	4.6 - 4.7	1.5-1.9	1.1-1.7	high
3	4.6 - 4.7	1.5-1.9	1.1-1.7	high

Table 8 Discretized Dataset with $Length_{4.4,4.6}$, $Height_{1.5}$ and $Weight_{1.1}$

$$\{A^d\}^* = \{\{5\},\{6\},\{2\},\{8\},\{1,3\},\{4\},\{7\}\}$$
$$\{d\}^* = \{\{1,2,3\},\{4,5\},\{6\},\{7,8\}\}$$
$$\{A\}^* \leq \{d\}^*$$

▶ Dataset is consistent.

Merging Redundant Cutpoints: In the discretized dataset, there are following cutpoints:

- Length: 4.4, 4.6
- Height: 1.5
- Weight: 1.1

Try Eliminating $Length_{4.4}$**:**

Cases	Attributes			Decision
	Length	Height	Weight	Quality
2	4.3 - 4.6	1.4-1.5	0.9-1.1	high
6	4.3 - 4.6	1.4-1.5	0.9-1.1	low
8	4.3 - 4.6	1.4-1.5	1.1-1.7	very-low
5	4.3 - 4.6	1.5-1.9	1.1-1.7	medium
4	4.3 - 4.6	1.5-1.9	1.1-1.7	medium
7	4.3 - 4.6	1.5-1.9	0.9-1.1	very-low
1	4.6 - 4.7	1.5-1.9	1.1-1.7	high
3	4.6 - 4.7	1.5-1.9	1.1-1.7	high

Table 9 Discretized Dataset after Eliminating $Length_{4.4}$

$$\{A^d\}^* = \{\{5,4\},\{2,6\},\{8\},\{1,3\},\{7\}\}$$
$$\{d\}^* = \{\{1,2,3\},\{4,5\},\{6\},\{7,8\}\}$$
$$\{A\}^* \not\leq \{d\}^*$$

▷ Dataset is inconsistent. Hence $Length_{4.4}$ cannot be eliminated.

Try Eliminating $Length_{4.6}$**:**

Cases	Attributes			Decision
	Length	Height	Weight	Quality
6	4.3 - 4.4	1.4-1.5	0.9-1.1	low
5	4.3 - 4.4	1.5-1.9	1.1-1.7	medium
2	4.4 - 4.7	1.4-1.5	0.9-1.1	high
8	4.4 - 4.7	1.4-1.5	1.1-1.7	very-low
7	4.4 - 4.7	1.5-1.9	0.9-1.1	very-low
1	4.4 - 4.7	1.5-1.9	1.1-1.7	high
3	4.4 - 4.7	1.5-1.9	1.1-1.7	high
4	4.4 - 4.7	1.5-1.9	1.1-1.7	medium

Table 10 Discretized Dataset after Eliminating $Length_{4.6}$

$$\{A^d\}^* = \{\{5\},\{2\},6\},\{8\},\{1,3,4\},\{7\}\}$$
$$\{d\}^* = \{\{1,2,3\},\{4,5\},\{6\},\{7,8\}\}$$
$$\{A\}^* \not\leq \{d\}^*$$

▷ Dataset is inconsistent. Hence $Length_{4.6}$ cannot be eliminated.

Try Eliminating $Height_{1.5}$**:**

Cases	Attributes			Decision
	Length	Height	Weight	Quality
6	4.3 - 4.4	1.4-1.9	0.9-1.1	low
5	4.3 - 4.4	1.4-1.9	1.1-1.7	medium
2	4.4 - 4.6	1.4-1.9	0.9-1.1	high
7	4.4 - 4.6	1.4-1.9	0.9-1.1	very-low
4	4.4 - 4.6	1.4-1.9	1.1-1.7	medium
8	4.4 - 4.6	1.4-1.9	1.1-1.7	very-low
1	4.6 - 4.7	1.4-1.9	1.1-1.7	high
3	4.6 - 4.7	1.4-1.9	1.1-1.7	high

Table 11 Discretized Dataset after Eliminating $Height_{1.5}$

$$\{A^d\}^* = \{\{5\},\{6\},\{2,7\},\{4,8\},\{1,3\}\}$$
$$\{d\}^* = \{\{1,2,3\},\{4,5\},\{6\},\{7,8\}\}$$
$$\{A\}^* \not\leq \{d\}^*$$

▷ Dataset is inconsistent. Hence $Height_{1.5}$ cannot be eliminated.

Try Eliminating $Weight_{1.1}$**:**

Cases	Attributes			Decision
	Length	Height	Weight	Quality
6	4.3 - 4.4	1.4-1.5	0.9-1.7	low
5	4.3 - 4.4	1.5-1.9	0.9-1.7	medium
2	4.4 - 4.6	1.4-1.5	0.9-1.7	high
8	4.4 - 4.6	1.4-1.5	0.9-1.7	very-low
4	4.4 - 4.6	1.5-1.9	0.9-1.7	medium
7	4.4 - 4.6	1.5-1.9	0.9-1.7	very-low
1	4.6 - 4.7	1.5-1.9	0.9-1.7	high
3	4.6 - 4.7	1.5-1.9	0.9-1.7	high

Table 12 Discretized Dataset after Eliminating $Weight_{1.1}$

$$\{A^d\}^* = \{\{5\},\{6\},\{2,8\},\{4,7\},\{1,3\}\}$$
$$\{d\}^* = \{\{1,2,3\},\{4,5\},\{6\},\{7,8\}\}$$
$$\{A\}^* \not\leq \{d\}^*$$

▷ Dataset is inconsistent. Hence $Weight_{1.1}$ cannot be eliminated.

▶ None of the cutpoints can be eliminated!

Part-(b) - Multiple Scanning Approach

Cases	Attributes				Decision
	Length	Height	Width	Weight	Quality
1	4.7	1.8	1.7	1.7	high
2	4.5	1.4	1.8	0.9	high
3	4.7	1.8	1.9	1.3	high
4	4.5	1.8	1.7	1.3	medium
5	4.3	1.6	1.9	1.7	medium
6	4.3	1.4	1.7	0.9	low
7	4.5	1.6	1.9	0.9	very-low
8	4.5	1.4	1.8	1.3	very-low

Table 13 Dataset

Discretize Length:
Cutpoints for Length: 4.4, 4.6

Trying $Length_{4.4}$**:**

$$\begin{aligned}
Length_{4.3-4.4} &= Cases(5,6) = (medium, low) \\
Length_{4.4-4.7} &= Cases(1,2,3,4,7,8) = (high, high, high, medium, very-low, very-low) \\
H(Quality|Length_{4.4}) &= -\frac{2}{8}(-1) - \frac{6}{8}(\frac{3}{6}log_2(\frac{3}{6}) + \frac{1}{6}log_2(\frac{1}{6}) + \frac{2}{6}log_2(\frac{2}{6})) \\
&= 1.344
\end{aligned}$$

Trying $Length_{4.6}$**:**

$$\begin{aligned}
Length_{4.3-4.6} &= Cases(2,4,5,6,7,8) = (high, medium, medium, low, very-low, very-low) \\
Length_{4.6-4.7} &= Cases(1,3) = (high, high) \\
H(Quality|Length_{4.6}) &= -\frac{6}{8}((\frac{1}{6}log_2(\frac{1}{6}))2 + (\frac{2}{6}log_2(\frac{2}{6}))2) - \frac{2}{8}(0) \\
&= 1.439
\end{aligned}$$

▷ Cutpoint 4.4 is better for *Length*.

Discretize Height:
Cutpoints for Height: 1.5, 1.7

Trying $Height_{1.5}$**:**

$$\begin{aligned}
Height_{1.4-1.5} &= Cases(2,6,8) = (high, low, very-low) \\
Height_{1.5-1.8} &= Cases(1,3,4,5,7) = (high, high, medium, medium, very-low) \\
H(Quality|Height_{1.5}) &= -\frac{3}{8}(log_2\frac{1}{3}) - \frac{5}{8}(\frac{4}{5}log_2(\frac{2}{5}) + \frac{1}{5}log_2(\frac{1}{5})) \\
&= 1.545
\end{aligned}$$

Trying $Height_{1.7}$**:**

$$\begin{aligned}
Height_{1.4-1.7} &= Cases(2,5,6,7,8) = (high, medium, low, very-low, very-low)\\
Height_{1.7-1.8} &= Cases(1,3,4) = (high, high, medium)\\
H(Quality|Height_{1.5}) &= -\frac{5}{8}(\frac{2}{5}log_2(\frac{2}{5}) + \frac{3}{5}log_2(\frac{1}{5})) - \frac{3}{8}(\frac{1}{3}log_2\frac{1}{3} + \frac{2}{3}log_2(\frac{2}{3}))\\
&= 1.545
\end{aligned}$$

▷ Both cutpoints for Height are equivalent. Picking $Height_{1.5}$.

Discretize Width:
Cutpoints for Width: 1.75, 1.85

Trying $Width_{1.75}$**:**

$$\begin{aligned}
Width_{1.7-1.75} &= Cases(1,4,6) = (high, medium, low)\\
Width_{1.75-1.9} &= Cases(2,3,5,7,8) = (high, high, medium, very-low, very-low)\\
H(Quality|Width_{1.75}) &= -\frac{3}{8}(log_2\frac{1}{3}) - \frac{5}{8}(\frac{4}{5}log_2(\frac{2}{5}) + \frac{1}{5}log_2(\frac{1}{5}))\\
&= 1.545
\end{aligned}$$

Trying $Width_{1.85}$**:**

$$\begin{aligned}
Width_{1.7-1.85} &= Cases(1,2,4,6,8) = (high, high, medium, low, very-low)\\
Width_{1.85-1.9} &= Cases(3,5,7) = (high, medium, very-low)\\
H(Quality|Width_{1.85}) &= -\frac{5}{8}(\frac{2}{5}log_2(\frac{2}{5}) + \frac{3}{5}log_2(\frac{1}{5})) - \frac{3}{8}(log_2\frac{1}{3})\\
&= 1.796
\end{aligned}$$

▷ Cutpoint 1.75 is better for $Width$.

Discretize Weight:
Cutpoints for Weight: 1.1, 1.5

Trying $Weight_{1.1}$**:**

$$\begin{aligned}
Weight_{0.9-1.1} &= Cases(2,6,7) = (high, low, very-low)\\
Weight_(1.1-1.7) &= Cases(1,3,4,5,8) = (high, high, medium, medium, very-low)\\
H(Quality|Weight_{1.1}) &= -\frac{3}{8}(log_2\frac{1}{3}) - \frac{5}{8}(\frac{4}{5}log_2(\frac{2}{5}) + \frac{1}{5}log_2(\frac{1}{5}))\\
&= 1.545
\end{aligned}$$

Trying $Weight_{1.5}$**:**

$$\begin{aligned}
Weight_{0.9-1.5} &= Cases(2,3,4,6,7,8) = (high, high, medium, low, very-low, very-low)\\
Weight_(1.5-1.7) &= Cases(1,5) = (high, medium)\\
H(Quality|Weight_{1.5}) &= -\frac{2}{8}(-1) - \frac{6}{8}(\frac{4}{6}log_2(\frac{2}{6}) + \frac{2}{6}log_2(\frac{1}{6}))\\
&= 1.689
\end{aligned}$$

▷ Cutpoint 1.1 is better for $Weight$.

Discretized Dataset after First Scan: In the discretized dataset, there are following cutpoints:

- Length: 4.4
- Height: 1.5
- Width: 1.75
- Weight: 1.1

Cases	Attributes				Decision
	Length	Height	Width	Weight	Quality
6	4.3-4.4	1.4-1.5	1.7-1.75	0.9-1.1	low
5	4.3-4.4	1.5-1.8	1.75-1.9	1.1-1.7	medium
2	4.4-4.7	1.4-1.5	1.75-1.9	0.9-1.1	high
8	4.4-4.7	1.4-1.5	1.75-1.9	1.1-1.7	very-low
1	4.4-4.7	1.5-1.8	1.7-1.75	1.1-1.7	high
4	4.4-4.7	1.5-1.8	1.7-1.75	1.1-1.7	medium
3	4.4-4.7	1.5-1.8	1.75-1.9	1.1-1.7	high
7	4.4-4.7	1.5-1.8	1.75-1.9	0.9-1.1	very-low

Table 14 Discretized Dataset after First Scan

$$\{A^d\}^* = \{\{1,4\},\{2\},\{3\},\{5\},\{6\},\{7\},\{8\}\}$$
$$\{d\}^* = \{1,2,3\},\{4,5\},\{6\},\{7,8\}\}$$
$$\{A^d\}^* \not\leq \{d\}^*$$
$$L(A) = \frac{2+1+1+2}{8} = \frac{6}{8} < 1$$

▷ Dataset is inconsistent!

Cases	Attributes				Decision
	Length	Height	Width	Weight	Quality
1	4.7	1.8	1.7	1.7	high
4	4.5	1.8	1.7	1.3	medium

Table 15 Conflicting Cases after First Scan

▷ Additional Cutpoint: $Length_{4.6}$

Cases	Attributes				Decision
	Length	Height	Width	Weight	Quality
6	4.3-4.4	1.4-1.5	1.7-1.75	0.9-1.1	low
5	4.3-4.4	1.5-1.8	1.75-1.9	1.1-1.7	medium
2	4.4-4.6	1.4-1.5	1.75-1.9	0.9-1.1	high
8	4.4-4.6	1.4-1.5	1.75-1.9	1.1-1.7	very-low
4	4.4-4.6	1.5-1.8	1.7-1.75	1.1-1.7	medium
7	4.4-4.6	1.5-1.8	1.75-1.9	0.9-1.1	very-low
1	4.6-4.7	1.5-1.8	1.7-1.75	1.1-1.7	high
3	4.6-4.7	1.5-1.8	1.75-1.9	1.1-1.7	high

Table 16 Discretized Dataset

$$\{A^d\}^* = \{\{1\},\{2\},\{3\},\{4\},\{5\},\{6\},\{7\},\{8\}\}$$
$$\{d\}^* = \{1,2,3\},\{4,5\},\{6\},\{7,8\}\}$$
$$\{A^d\}^* \leq \{d\}^*$$
$$L(A) = \frac{3+2+1+2}{8} = \frac{8}{8} = 1$$

▶ Dataset is consistent!

Merging Redundant Cutpoints: In the discretized dataset, there are following cutpoints:

- Length: 4.4, 4.6
- Height: 1.5
- Width: 1.75
- Weight: 1.1

Try Eliminating $Length_{4.4}$**:**

Cases	Attributes				Decision
	Length	Height	Width	Weight	Quality
6	4.3-4.6	1.4-1.5	1.7-1.75	0.9-1.1	low
2	4.3-4.6	1.4-1.5	1.75-1.9	0.9-1.1	high
8	4.3-4.6	1.4-1.5	1.75-1.9	1.1-1.7	very-low
4	4.3-4.6	1.5-1.8	1.7-1.75	1.1-1.7	medium
7	4.3-4.6	1.5-1.8	1.75-1.9	0.9-1.1	very-low
5	4.3-4.6	1.5-1.8	1.75-1.9	1.1-1.7	medium
1	4.6-4.7	1.5-1.8	1.7-1.75	1.1-1.7	high
3	4.6-4.7	1.5-1.8	1.75-1.9	1.1-1.7	high

Table 17 Discretized Dataset after Eliminating $Length_{4.4}$

$$\{A^d\}^* = \{\{1\},\{2\},\{3\},\{4\},\{5\},\{6\},\{7\},\{8\}\}$$
$$\{d\}^* = \{1,2,3\},\{4,5\},\{6\},\{7,8\}\}$$
$$\{A^d\}^* \leq \{d\}^*$$

► Dataset is consistent. Hence $Length_{4.4}$ can be eliminated!

▷ $Length_{4.6}$ cannot be eliminated since it was added in the last step of discretization to distinguish between cases 1 and 4.

Try Eliminating $Height_{1.5}$**:**

Cases	Attributes				Decision
	Length	Height	Width	Weight	Quality
6	4.3-4.6	1.4-1.5	1.7-1.75	0.9-1.1	low
4	4.3-4.6	1.4-1.8	1.7-1.75	1.1-1.7	medium
2	4.3-4.6	1.4-1.5	1.75-1.9	0.9-1.1	high
7	4.3-4.6	1.4-1.8	1.75-1.9	0.9-1.1	very-low
5	4.3-4.6	1.4-1.8	1.75-1.9	1.1-1.7	medium
8	4.3-4.6	1.4-1.5	1.75-1.9	1.1-1.7	very-low
1	4.6-4.7	1.4-1.8	1.7-1.75	1.1-1.7	high
3	4.6-4.7	1.4-1.8	1.75-1.9	1.1-1.7	high

Table 18 Discretized Dataset after Eliminating $Height_{1.5}$

$$\{A^d\}^* = \{\{1\}, \{2,7\}, \{3\}, \{4\}, \{5,8\}, \{6\}\}$$
$$\{d\}^* = \{1,2,3\}, \{4,5\}, \{6\}, \{7,8\}\}$$
$$\{A^d\}^* \nleq \{d\}^*$$

▷ $Height_{1.5}$ cannot be eliminated.

Try Eliminating $Width_{1.75}$:

Cases	Attributes				Decision
	Length	Height	Width	Weight	Quality
6	4.3-4.6	1.4-1.5	1.7-1.9	0.9-1.1	low
2	4.3-4.6	1.4-1.5	1.7-1.9	0.9-1.1	high
8	4.3-4.6	1.4-1.5	1.7-1.9	1.1-1.7	very-low
7	4.3-4.6	1.5-1.8	1.7-1.9	0.9-1.1	very-low
4	4.3-4.6	1.5-1.8	1.7-1.9	1.1-1.7	medium
5	4.3-4.6	1.5-1.8	1.7-1.9	1.1-1.7	medium
1	4.6-4.7	1.5-1.8	1.7-1.9	1.1-1.7	high
3	4.6-4.7	1.5-1.8	1.7-1.9	1.1-1.7	high

Table 19 Discretized Dataset after Eliminating $Width_{1.75}$

$$\{A^d\}^* = \{\{1,3\}, \{2,6\}, \{4,5\}, \{7\}, \{8\}\}$$
$$\{d\}^* = \{1,2,3\}, \{4,5\}, \{6\}, \{7,8\}\}$$
$$\{A^d\}^* \nleq \{d\}^*$$

▷ $Width1.75$ cannot be eliminated.

Try Eliminating $Weight1.1$:

Cases	Attributes				Decision
	Length	Height	Width	Weight	Quality
6	4.3-4.6	1.4-1.5	1.7-1.75	0.9-1.7	low
2	4.3-4.6	1.4-1.5	1.75-1.9	0.9-1.7	high
8	4.3-4.6	1.4-1.5	1.75-1.9	0.9-1.7	very-low
4	4.3-4.6	1.5-1.8	1.7-1.75	0.9-1.7	medium
7	4.3-4.6	1.5-1.8	1.75-1.9	0.9-1.7	very-low
5	4.3-4.6	1.5-1.8	1.75-1.9	0.9-1.7	medium
1	4.6-4.7	1.5-1.8	1.7-1.75	0.9-1.7	high
3	4.6-4.7	1.5-1.8	1.75-1.9	0.9-1.7	high

Table 20 Discretized Dataset after Eliminating $Weight_{1.1}$

$$\{A^d\}^* = \{\{1\},\{2,8\},\{3\},\{4\},\{5,7\},\{6\}\}$$
$$\{d\}^* = \{1,2,3\},\{4,5\},\{6\},\{7,8\}\}$$
$$\{A^d\}^* \not\leq \{d\}^*$$

▷ $Weight1.1$ cannot be eliminated.

Final Discretized Dataset:

Cases	Attributes				Decision
	Length	Height	Width	Weight	Quality
1	4.6-4.7	1.5-1.8	1.7-1.75	1.1-1.7	high
2	4.3-4.6	1.4-1.5	1.75-1.9	0.9-1.1	high
3	4.6-4.7	1.5-1.8	1.75-1.9	1.1-1.7	high
4	4.3-4.6	1.5-1.8	1.7-1.75	1.1-1.7	medium
5	4.3-4.6	1.5-1.8	1.75-1.9	1.1-1.7	medium
6	4.3-4.6	1.4-1.5	1.7-1.75	0.9-1.1	low
7	4.3-4.6	1.5-1.8	1.75-1.9	0.9-1.1	very-low
8	4.3-4.6	1.4-1.5	1.75-1.9	1.1-1.7	very-low

Table 21 Final Discretized Dataset

Part-(c): Globalized Equal Frequency per Interval Approach

First of all, we discretize the dataset with k = 2:

Cases	Attributes				Decision
	Length	Height	Width	Weight	Quality
1	4.6-4.7	1.7-1.8	1.7-1.85	1.1-1.7	high
2	4.3-4.6	1.4-1.7	1.7-1.85	0.9-1.1	high
3	4.6-4.7	1.7-1.8	1.85-1.9	1.1-1.7	high
4	4.3-4.6	1.7-1.8	1.7-1.85	1.1-1.7	medium
5	4.3-4.6	1.4-1.7	1.85-1.9	1.1-1.7	medium
6	4.3-4.6	1.4-1.7	1.7-1.85	0.9-1.1	low
7	4.3-4.6	1.4-1.7	1.85-1.9	0.9-1.1	very-low
8	4.3-4.6	1.4-1.7	1.7-1.85	1.1-1.7	very-low

Table 22 Dataset Discretized via Equal Frequency per Interval (k = 2)

Level of Consistency:

$$\{A^d\}^* = \{\{1\},\{2,6\},\{3\},\{4\},\{5\},\{7\},\{8\}\}$$
$$\{d\}^* = \{\{1,2,3\},\{4,5\},\{6\},\{7,8\}\}$$
$$L(A) = \frac{2+2+2}{8} = \frac{6}{8} < 1$$

▷ Dataset is not consistent.

Average Block Entropy of $Length^d$:

$$\{Length^d\}^* = \{\{2,4,5,6,7,8\},\{1,3\}\}$$
$$M(Length^d) = \frac{\frac{6}{8}(-\frac{4}{6}log_2(\frac{2}{6}) - \frac{2}{6}log_2(\frac{1}{6})) + \frac{2}{8}(0)}{2}$$
$$= 0.720$$

Average Block Entropy of $Height^d$:

$$\{Height^d\}^* = \{\{2,5,6,7,8\},\{1,3,4\}\}$$
$$M(Height^d) = \frac{\frac{5}{8}(-\frac{3}{5}log_2(\frac{1}{5}) - \frac{2}{5}log_2(\frac{2}{5})) + \frac{3}{8}(-\frac{2}{3}log_2(\frac{2}{3}) - \frac{1}{3}log_2(\frac{1}{3}))}{2}$$
$$= 0.773$$

Average Block Entropy of $Width^d$:

$$
\begin{aligned}
\{Width^d\}^* &= \{\{1,2,4,6,8\},\{3,5,7\}\} \\
M(Width^d) &= \frac{\frac{5}{8}(-\frac{2}{5}log_2(\frac{2}{5}) - \frac{3}{5}log_2(\frac{1}{5})) + \frac{3}{8}(-log_2(\frac{1}{3}))}{2} \\
&= 0.898
\end{aligned}
$$

Average Block Entropy of $Weight^d$:

$$
\begin{aligned}
\{Weight^d\}^* &= \{\{2,6,7\},\{1,3,4,5,8\}\} \\
M(Weight^d) &= \frac{\frac{5}{8}(-\frac{4}{5}log_2(\frac{2}{5}) - \frac{1}{5}log_2(\frac{1}{5})) + \frac{3}{8}(-log_2(\frac{1}{3}))}{2} \\
&= 0.773
\end{aligned}
$$

Worst Attribute: Width

Cases	Attributes				Decision
	Length	Height	Width	Weight	Quality
1	4.6-4.7	1.7-1.8	1.7-1.75	1.1-1.7	high
2	4.3-4.6	1.4-1.7	1.75-1.85	0.9-1.1	high
3	4.6-4.7	1.7-1.8	1.85-1.9	1.1-1.7	high
4	4.3-4.6	1.7-1.8	1.7-1.75	1.1-1.7	medium
5	4.3-4.6	1.4-1.7	1.85-1.9	1.1-1.7	medium
6	4.3-4.6	1.4-1.7	1.7-1.75	0.9-1.1	low
7	4.3-4.6	1.4-1.7	1.85-1.9	0.9-1.1	very-low
8	4.3-4.6	1.4-1.7	1.75-1.85	1.1-1.7	very-low

Table 23 Further Discretized Dataset with $k = 3$ for Width

Level of Consistency:

$$
\begin{aligned}
\{A^d\}^* &= \{\{1\},\{2\},\{3\},\{4\},\{5\},\{6\},\{7\},\{8\}\} \\
\{d\}^* &= \{\{1,2,3\},\{4,5\},\{6\},\{7,8\}\} \\
L(A) &= \frac{3+2++1+2}{8} = \frac{8}{8} = 1
\end{aligned}
$$

▷ Dataset is consistent!

Merging Redundant Cutpoints:
The dataset has the following cutpoints:

- Length: 4.6
- Height: 1.7
- Width: 1.75, 1.85
- Weight: 1.1

Cases	Attributes				Decision
	Length	Height	Width	Weight	Quality
6	4.3-4.6	1.4-1.7	1.7-1.75	0.9-1.1	low
2	4.3-4.6	1.4-1.7	1.75-1.85	0.9-1.1	high
8	4.3-4.6	1.4-1.7	1.75-1.85	1.1-1.7	very-low
7	4.3-4.6	1.4-1.7	1.85-1.9	0.9-1.1	very-low
5	4.3-4.6	1.4-1.7	1.85-1.9	1.1-1.7	medium
4	4.3-4.6	1.7-1.8	1.7-1.75	1.1-1.7	medium
1	4.6-4.7	1.7-1.8	1.7-1.75	1.1-1.7	high
3	4.6-4.7	1.7-1.8	1.85-1.9	1.1-1.7	high

Table 24 Discretized Dataset with Cutpoints

Try Eliminating $Length_{4.6}$**:**

$$\{A^d\}^* = \{\{1,4\},\{2\},\{3\},\{5\},\{6\},\{7\},\{8\}\}$$
$$\{d\}^* = \{\{1,2,3\},\{4,5\},\{6\},\{7,8\}\}$$
$$L(A) = \frac{2+1++1+2}{8} = \frac{6}{8} < 1$$

▷ Dataset is not consistent. Hence $Length_{4.6}$ cannot be eliminated.

Try Eliminating $Height_{1.7}$**:**

$$\{A^d\}^* = \{\{1\},\{2\},\{3\},\{4\},\{5\},\{6\},\{7\},\{8\}\}$$
$$\{d\}^* = \{\{1,2,3\},\{4,5\},\{6\},\{7,8\}\}$$
$$L(A) = \frac{3+2++1+2}{8} = \frac{8}{8} = 1$$

► Dataset is consistent. Hence $Height_{1.7}$ cann be eliminated!

Cases	Attributes			Decision
	Length	Width	Weight	Quality
6	4.3-4.6	1.7-1.75	0.9-1.1	low
4	4.3-4.6	1.7-1.75	1.1-1.7	medium
2	4.3-4.6	1.75-1.85	0.9-1.1	high
8	4.3-4.6	1.75-1.85	1.1-1.7	very-low
7	4.3-4.6	1.85-1.9	0.9-1.1	very-low
5	4.3-4.6	1.85-1.9	1.1-1.7	medium
1	4.6-4.7	1.7-1.75	1.1-1.7	high
3	4.6-4.7	1.85-1.9	1.1-1.7	high

Table 25 Discretized Dataset with Cutpoints after Eliminating $Height_{1.7}$

Try Eliminating $Width_{1.75}$**:**

$$\{A^d\}^* = \{\{1\},\{2,6\},\{3\},\{4\},\{5\},\{7\},\{8\}\}$$
$$\{d\}^* = \{\{1,2,3\},\{4,5\},\{6\},\{7,8\}\}$$
$$L(A) = \frac{2+2+2}{8} = \frac{6}{8} < 1$$

▷ Dataset is inconsistent. Hence $Width_{1.75}$ cannot be eliminated.

Try Eliminating $Width_{1.85}$**:**

$$\{A^d\}^* = \{\{1\},\{2,7\},\{3\},\{4\},\{5,8\}\}$$
$$\{d\}^* = \{\{1,2,3\},\{4,5\},\{6\},\{7,8\}\}$$
$$L(A) = \frac{2+1+1}{8} = \frac{4}{8} < 1$$

▷ Dataset is inconsistent. Hence $Width_{1.85}$ cannot be eliminated.

Try Eliminating $Weight_{1.1}$**:**

$$\{A^d\}^* = \{\{1\},\{2,8\},\{3\},\{4,6\},\{5,7\}\}$$
$$\{d\}^* = \{\{1,2,3\},\{4,5\},\{6\},\{7,8\}\}$$
$$L(A) = \frac{2}{8} = \frac{1}{4} < 1$$

▷ Dataset is inconsistent. Hence $Weight_{1.1}$ cannot be eliminated.

Final Discretized Dataset:

- Length: 4.6
- Width: 1.75, 1.85
- Weight: 1.1

Cases	Attributes			Decision
	Length	Width	Weight	Quality
1	4.6-4.7	1.7-1.75	1.1-1.7	high
2	4.3-4.6	1.75-1.85	0.9-1.1	high
3	4.6-4.7	1.85-1.9	1.1-1.7	high
4	4.3-4.6	1.7-1.75	1.1-1.7	medium
5	4.3-4.6	1.85-1.9	1.1-1.7	medium
6	4.3-4.6	1.7-1.75	0.9-1.1	low
7	4.3-4.6	1.85-1.9	0.9-1.1	very-low
8	4.3-4.6	1.75-1.85	1.1-1.7	very-low

Table 26 Final Discretized Dataset