1. **Problem 1**

Attribute	Values
Chest Pain(CPT)	AP, NA, NP
Fasting Blood Sugar(FBS)	$1, 0 > 120mg$
Max Heart Rate Achieved(MHRA)	L, A, H
Number major vessels(NMV)	$0, 1, 2, 3$
Heart Disease(HD)	$0, 1$

   (a) Iteration 1:

   Information Before:

   $$8+, 8-$$

   $$H\left(\frac{1}{2}\right) = -\frac{1}{2}\log_2\frac{1}{2} - \frac{1}{2}\log_2\frac{1}{2} = 1$$

   Information After:

   i. CPT:

   $$
   \begin{array}{l|l}
   AP : 1+, 3- & H\left(\frac{1}{4}\right) = -\frac{1}{4}\log_2\frac{1}{4} - \frac{3}{4}\log_2\frac{3}{4} \approx 0.811278 \\
   NA : 1+, 4- & H\left(\frac{1}{5}\right) = -\frac{1}{5}\log_2\frac{1}{5} - \frac{4}{5}\log_2\frac{4}{5} \approx 0.721928 \\
   NP : 6+, 1- & H\left(\frac{6}{7}\right) = -\frac{6}{7}\log_2\frac{6}{7} - \frac{1}{7}\log_2\frac{1}{7} \approx 0.5916728
   \end{array}
   $$

   $$\frac{1}{4} \cdot 0.811278 + \frac{5}{16} \cdot 0.721928 + \frac{7}{16} \cdot 0.5916728 \approx 0.6872789$$

   $$Gain(CPT) = 1 - 0.6872789 \approx 0.31272115$$

   ii. FBS:

   $$
   \begin{array}{l|l}
   0 : 2+, 1- & H\left(\frac{2}{3}\right) = -\frac{2}{3}\log_2\frac{2}{3} - \frac{1}{3}\log_2\frac{1}{3} \approx 0.918296 \\
   1 : 6+, 1- & H\left(\frac{6}{7}\right) = -\frac{6}{7}\log_2\frac{6}{7} - \frac{1}{7}\log_2\frac{1}{7} \approx 0.5916728
   \end{array}
   $$

   $$\frac{3}{16} \cdot 0.918296 + \frac{13}{16} \cdot 0.5916728 \approx 0.65291465$$

   $$Gain(FBS) = 1 - 0.65291465 \approx 0.34708535$$

iii. MHRA:

$$\begin{array}{l|l}
L: 4+, 1- & H(\frac{1}{5}) = -\frac{1}{5}\log_2\frac{1}{5} - \frac{4}{5}\log_2\frac{4}{5} \approx 0.721928 \\
A: 3+, 5- & H(\frac{3}{8}) = -\frac{3}{8}\log_2\frac{3}{8} - \frac{5}{8}\log_2\frac{5}{8} \approx 0.954434 \\
H: 1+, 2- & H(\frac{1}{3}) = -\frac{1}{3}\log_2\frac{1}{3} - \frac{2}{3}\log_2\frac{2}{3} \approx 0.918295834
\end{array}$$

$$\frac{5}{16}\cdot 0.721928 + \frac{1}{2}\cdot 0.954434 + \frac{3}{16}\cdot 0.918295834 \approx 0.875$$

$$Gain(MHRA) = 1 - 0.875 \approx 0.125$$

iv. NMV:

$$\begin{array}{l|l}
0: 1+, 6- & H(\frac{1}{7}) = -\frac{1}{7}\log_2\frac{1}{7} - \frac{6}{7}\log_2\frac{6}{7} \approx 0.5916728 \\
1: 3+, 2- & H(\frac{3}{5}) = -\frac{3}{5}\log_2\frac{3}{5} - \frac{2}{5}\log_2\frac{2}{5} \approx 0.970951 \\
2: 2+, 0- & H = 0 \\
3: 2+, 0- & H = 0
\end{array}$$

$$\frac{7}{16}\cdot 0.591673 + \frac{5}{16}\cdot 0.970951 + 0 + 0 \approx 0.5622789$$

$$Gain(MHRA) = 1 - 0.5622789 \approx 0.4377211$$

Choose NVM

Resultant Decision Tree:

```
1       if <NMV> = 0  :
2            <class> = 0
3       if <NMV> = 1  :
4            <class> = 1
5       if <NMV> = 2  :
6            <class> = 1
7       if <NMV> = 3  :
8            <class> = 1
```

(b) Iteration 2:

Information Before:

$$H = \left\{ \begin{array}{ll} 0 & \to 0.591673 \\ 1 & \to 0.970951 \end{array} \right\}$$

Information After:

i. CPT:

A. NMV_0

$$\begin{array}{l|l}
AP: 0+, 2- & H = 0 \\
NA: 0+, 3- & H = 0 \\
NP: 1+, 1- & H = 1
\end{array}$$

$$Gain(CPT) = 0.591673 - \frac{2}{7} \approx 0.3059585$$

B. NMV_1

$$
\begin{array}{ll|l}
AP : 1+, 1- & H = 1 \\
NA : 1+, 1- & H = 1 \\
NP : 1+, 0- & H = 0
\end{array}
$$

$$Gain(CPT) = 0.970951 - \frac{2}{5} - \frac{2}{5} \approx 0.17095$$

ii. FBS:

A. NMV_0

$$
\begin{array}{ll|l}
0 : 0+, 5- & H = 0 \\
1 : 1+, 1- & H = 1
\end{array}
$$

$$Gain(FBS) = 0.591673 - \frac{2}{7} \approx 0.3059585$$

B. NMV_1

$$
\begin{array}{ll|l}
0 : 3+, 2- & H = -\frac{3}{5}\log_2\frac{3}{5} - \frac{2}{5}\log_2\frac{2}{5} \approx 0.9709505945 \\
1 : 1+, 1- & H = 1
\end{array}
$$

$$Gain(FBS) = 0.970951 - 0.9709505945 \approx 0.0$$

iii. MHRA:

A. NMV_0

$$
\begin{array}{ll|l}
L : 0+, 1- & H = 0 \\
A : 1+, 3- & H \approx 0.81127812 \\
H : 0+, 2- & H = 0
\end{array}
$$

$$Gain(MHRA) = 0.591673 - \frac{4}{7} \cdot 0.81127812 \approx 0.1280853$$

B. NMV_1

$$
\begin{array}{ll|l}
L : 2+, 0- & H = 0 \\
A : 1+, 2- & H \approx 0.9182958 \\
H : 0+, 0- & H = 0
\end{array}
$$

$$Gain(MHRA) = 0.970951 - \frac{3}{5} \cdot 0.9182958 \approx 0.41997311$$

Choose MHRA

Resultant Decision Tree:

```
1            if <NMV> = 0 :
2              <class> = 0
3            if <NMV> = 1 :
4              if <MHRA> = L :
5                <class> = 1
6              if <MHRA> = A :
7                <class> = 0
8              if <MHRA> = H :
```

```
 9              <class> = 1
10         if <NMV> = 2 :
11            <class> = 1
12         if <NMV> = 3 :
13            <class> = 1
```

(c) Iteration 3:

Information Before:

  i. NMV_0

$$H = 0.591673$$

  ii. $MHRA$

$$H \approx 0.9182958$$

Information After:

  i. NMV_0

$$Gain(CPT) \approx 0.3059585$$
$$Gain(FBS) \approx 0.3059585$$

  ii. $MHRA$

    A. CPT:

$$
\begin{array}{ll|l}
AP: 0+, 1- & & H = 0 \\
NA: 0+, 1- & & H = 0 \\
NP: 1+, 0- & & H = 0
\end{array}
$$

$$Gain(CPT) \approx 0.9182958$$

    B. FBS:

$$
\begin{array}{ll|l}
0: 1+, 2- & & H = 0.918295834 \\
1: 0+, 0- & & H = 0
\end{array}
$$

$$Gain(CPT) = 0.0$$

Choose CPT

Resultant Decision Tree:

```
1          if <NMV> = 0 :
2            if <FBS> = 0 :
3              <class> = 0
4            if <FBS> = 1 :
5              <class> = 0
6          if <NMV> = 1 :
7            if <MHRA> = L :
8              <class> = 1
```

```
 9            if <MHRA> = A :
10              if <CPT> = AP :
11                <class> = 0
12              if <CPT> = NA :
13                <class> = 0
14              if <CPT> = NP :
15                <class> = 1
16            if <MHRA> = H :
17              <class> = 1
18          if <NMV> = 2 :
19            <class> = 1
20          if <NMV> = 3 :
21            <class> = 1
```

Accuracy on Table 2:

Index	Classification
1	1 → False Positive
2	1
3	0
4	0
5	1
6	1 → False Positive
7	1
8	1

Confusion Matrix:

4	0
2	2

Accuracy: $\frac{6}{8} = 75\%$

2. **Problem 2**

   (a) Depth 1:

   Current Tree:

```
1          if <NMV> = 0 :
2            <class> = 0
3          if <NMV> = 1 :
4            <class> = 1
5          if <NMV> = 2 :
6            <class> = 1
7          if <NMV> = 3 :
8            <class> = 1
```

Index	Classification
1	1 → False Positive
2	1
3	1 → False Positive
4	0
5	1
6	1 → False Positive
7	1
8	1

Accuracy: 62.5%

Depth 2:

Current Tree:

```
1          if <NMV> = 0 :
2            <class> = 0
3          if <NMV> = 1 :
4            if <MHRA> = L :
5              <class> = 1
6            if <MHRA> = A :
7              <class> = 0
8            if <MHRA> = H :
9              <class> = 1
10         if <NMV> = 2 :
11           <class> = 1
12         if <NMV> = 3 :
13           <class> = 1
```

Index	Classification
1	0
2	1
3	0
4	0
5	1
6	1 → False Positive
7	1
8	1

Accuracy: 87.5%

Depth 3:

Current Tree:

```
1          if <NMV> = 0 :
2            <class> = 0
3          if <NMV> = 1 :
4            if <MHRA> = L :
5              <class> = 1
6            if <MHRA> = A :
```

```
 7              if <CPT> = AP :
 8                 <class> = 0
 9              if <CPT> = NA :
10                 <class> = 0
11              if <CPT> = NP :
12                 <class> = 1
13           if <MHRA> = H :
14              <class> = 1
15        if <NMV> = 2 :
16           <class> = 1
17        if <NMV> = 3 :
18           <class> = 1
```

Index	Classification
1	1 → False Positive
2	1
3	0
4	0
5	1
6	1 → False Positive
7	1
8	1

Accuracy: 75%

(b) Reduced Error Pruning

    i. FBS node

Index	Classification	Classification After Pruning
1	0	0
2	0	0
3	0 → false negative	0 → false negative
4	1	1
5	0	0

Prune leaf with class 0

    ii. CPT node

Index	Classification	Classification After Pruning
1	0	0
2	0	0
3	0 → false negative	0 → false negative
4	1	1
5	0	0

Prune leaf with class 0

iii. MHRA node

Index	Classification	Classification After Pruning
1	0	0
2	0	0
3	0 → false negative	0 → false negative
4	1	0 → false negative
5	0	1 → false positive

Do not prune leaf

iv. NMV node

Index	Classification	Classification After Pruning
1	0	0
2	0	0
3	0 → false negative	0 → false negative
4	1	1
5	0	1 → false positive

Do not prune leaf

Resultant Tree:

```
 1          if <NMV> = 0 :
 2            <class> = 0
 3          if <NMV> = 1 :
 4            if <MHRA> = L :
 5              <class> = 1
 6            if <MHRA> = A :
 7              <class> = 0
 8            if <MHRA> = H :
 9              <class> = 1
10          if <NMV> = 2 :
11            <class> = 1
12          if <NMV> = 3 :
13            <class> = 1
```

Accuracy on Table 2:

Index	Classification
1	0
2	1
3	0
4	0
5	1
6	1 → False Positive
7	1
8	1

Confusion Matrix:

4	0
1	3

Accuracy: 87.5%

3. **Problem 3**

Information Before:

$$10+, 10-$$

$$H(\frac{1}{2}) = 1$$

(a) Iteration 1

Information After

i. CPT:

Let example 20 have a value of "NP" based on the following frequencies:
NA: 5, NP: 9, AP: 5

$$
\begin{array}{l|l}
AP : 1+, 4- & H \approx 0.721928 \\
NA : 1+, 4- & H \approx 0.721928 \\
NP : 8+, 2- & H \approx 0.721928
\end{array}
$$

$$Gain(CPT) \approx 0.278072$$

ii. FBS:

Let example 17 have a value of "0" based on the following frequencies:
0: 14, 1: 5

$$
\begin{array}{l|l}
0 : 2+, 2- & H \approx 1 \\
1 : 6+, 7- & H \approx 0.9957275
\end{array}
$$

$$Gain(FBS) \approx 0.152777125$$

iii. MHRA:

Let examples 18 and 19 have a value of "A" based on the following frequencies:
L: 5, A: 6, H: 5

$$
\begin{array}{l|l}
L : 4+, 1- & H \approx 0.721928 \\
A : 4+, 6- & H \approx 0.970951 \\
H : 2+, 3- & H \approx 0.970951
\end{array}
$$

$$Gain(MHRA) \approx 0.278072$$

iv. NMV:

Let examples 18 and 19 have a value of "A" based on the following frequencies:

L: 5, A: 6, H: 5

$$
\begin{array}{ll|l}
0: 1+, 7- & H \approx 0.54356444 \\
1: 3+, 2- & H \approx 0.970951 \\
2: 3+, 1- & H \approx 0.811278 \\
3: 3+, 0- & H = 0
\end{array}
$$

$$Gain(MHRA) \approx 0.377580874$$

Choose NMV

Resultant Decision Tree:

```
1        if <NMV> = 0 :
2            <class> = 0
3        if <NMV> = 1 :
4            <class> = 1
5        if <NMV> = 2 :
6            <class> = 1
7        if <NMV> = 3 :
8            <class> = 1
```

Information Before:

$$1+, 8-$$

$$H \approx 0.543564$$

(a) Iteration 1

Information After

i. NMV_0

A. CPT:

$$
\begin{array}{ll|l}
AP: 0+, 2- & H = 0 \\
NA: 0+, 3- & H = 0 \\
NP: 1+, 2- & H = 0.918295834
\end{array}
$$

$$Gain(CBT) \approx 0.19920344$$

B. FBS:

Let examples 17 have a value of "0" based on the following frequencies:

0: 5, 1: 2

$$
\begin{array}{ll|l}
0: 0+, 6- & H = 0 \\
1: 1+, 1- & H = 1
\end{array}
$$

$$Gain(FBS) \approx 0.293564$$

C. MHRA:

$$
\begin{array}{ll|l}
L : 0+, 1- & H = 0 \\
A : 1+, 3- & H = 0.81127812 \\
H : 0+, 3- & H = 0
\end{array}
$$

$$Gain(MHRA) \approx 0.137925381$$

ii. NMV_1

A. CPT:

$$
\begin{array}{ll|l}
AP : 1+, 1- & H = 1 \\
NA : 1+, 1- & H = 1 \\
NP : 1+, 0- & H = 0
\end{array}
$$

$$Gain(CBT) \approx 0.17095$$

B. FBS:

$$
\begin{array}{ll|l}
0 : 3+, 2- & H = 0.970951 \\
1 : 0+, 0- & H = 0
\end{array}
$$

$$Gain(FBS) \approx 0.0$$

C. MHRA:

$$
\begin{array}{ll|l}
L : 2+, 0- & H = 0 \\
A : 1+, 2- & H = 0.9182958 \\
H : 0+, 0- & H = 0
\end{array}
$$

$$Gain(MHRA) \approx 0.41997252$$

iii. NMV_2

A. CPT:
Let example 20 have a value of "NP" based on the following frequencies:
AP: 1, NA: 0, NP: 2

$$
\begin{array}{ll|l}
AP : 0+, 1- & H = 0 \\
NA : 0+, 0- & H = 0 \\
NP : 3+, 0- & H = 0
\end{array}
$$

$$Gain(CBT) \approx 0.97095$$

Since this has the maximum gain possible for this branch we will choose CPT to attach to NMV_2
This will result in the following tree:

```
1       if <NMV> = 0 :
2         <class> = 0
3       if <NMV> = 1 :
4         if <MHRA> = L :
5           <class> = 1
6         if <MHRA> = A :
7           <class> = 0
8         if <MHRA> = H :
9           <class> = 1
10      if <NMV> = 2 :
11        if <CPT> = AP :
12          <class> = 0
13        if <CPT> = NA :
14          <class> = 1
15        if <CPT> = NP :
16          <class> = 1
17      if <NMV> = 3 :
18        <class> = 1
```

i. $MHRA_A$
   Before Information

$$H = 0.9182958$$

   After Information
   A. FBS:

$$\begin{array}{l|l} 0 : 1+, 2- & H = 0.9182958 \\ 1 : 0+, 0- & H = 0 \end{array}$$

$$Gain(FBS) \approx 0.0$$

ii. NMV_0
   Before Information

$$H = 0.543564$$

   After Information
   A. FBS:
      Let examples 17 have a value of "0" based on the following frequencies:
      0: 5, 1: 2

$$\begin{array}{l|l} 0 : 0+, 6- & H = 0 \\ 1 : 1+, 1- & H = 1 \end{array}$$

$$Gain(FBS) \approx 0.293564$$

Choose FBS at NMV_0 which gives the following tree:

```
1               if <NMV> = 0 :
2                 if <FBS> = 0 :
3                    <class> = 0
4                 if <FBS> = 1 :
5                    <class> = 0
6               if <NMV> = 1 :
7                 if <MHRA> = L :
8                    <class> = 1
9                 if <MHRA> = A :
10                   <class> = 0
11                if <MHRA> = H :
12                   <class> = 1
13              if <NMV> = 2 :
14                if <CPT> = AP :
15                   <class> = 0
16                if <CPT> = NA :
17                   <class> = 1
18                if <CPT> = NP :
19                   <class> = 1
20              if <NMV> = 3 :
21                   <class> = 1
```