

**Faculté des Nouvelles Technologies de l'Information et de la Communication**

**Département : Informatique Fondamentale et ses Applications**

**Année Universitaire : 2023/2024**

**Module : DAMI (Master 1 SDIA)**

**Solution TD N° 3.2 : Arbres de décision**

**A. Solution avec l'algorithme ID3**

L'algorithme commence par évaluer chaque attribut disponible pour déterminer lequel serait le meilleur choix pour diviser les données. Pour ce faire, nous examinons les différentes valeurs de chaque attribut par rapport aux deux classes cibles, c'est-à-dire "Opérer" ou "Ne pas opérer". Le tableau suivant présente le nombre de cas "Oui" et "Non" pour chaque combinaison de valeurs d'attributs :

Age	Genre	Hypertension	Température	ECG	Hémoglobine	Glucose	Opérer ?
Vieux = 9	Masculin = 7	Normale = 5	Normale = 9	Normale = 9	Normale = 11	Normale = 5	Oui
Vieux = 4	Masculin = 1	Normale = 2	Normale = 4	Normale = 0	Normale = 1	Normale = 2	Non
Jeune = 2	Féminin = 4	Grave = 6	Grave = 2	Grave = 2	Grave = 0	Grave = 6	Oui
Jeune = 0	Féminin = 3	Grave = 2	Grave = 0	Grave = 4	Grave = 3	Grave = 2	Non

$$\text{Entropie : } H(X) = - \sum_{i=1}^n P_i \log_2(P_i)$$

$$\text{Gain : } \text{Gain}(X, a_i) = H(X) - \sum \frac{|X_{ai=v}|}{|X|} H(X_{ai=v})$$

Notre base de données contient 11 personnes qui ont opéré, et 4 personnes qui n'ont pas opéré, alors l'entropie est :

$$H(X) = - \left( \frac{11}{15} \log_2 \frac{11}{15} + \frac{4}{15} \log_2 \frac{4}{15} \right) \approx 0,83664074 \dots$$

**NB :** Pour simplifier le calcul de log base 2 du 11/15, vous pouvez utiliser le changement de base du logarithme. Voici comment vous pouvez le faire :

$$\log_2 \left( \frac{11}{15} \right) = \log \left( \frac{11}{15} \right) / \log(2)$$

**Calcul du gain "Age":**

$$\text{Gain}(X, \text{Age}) = H(X) - \sum \frac{|X_{\text{Age}=v}|}{|X|} H(X_{\text{Age}=v})$$

$$H(X_{\text{Age}=Vieux}) = -\left(\frac{9}{13} \log_2 \frac{9}{13} + \frac{4}{13} \log_2 \frac{4}{13}\right) \approx 0,89049164 \dots$$

$$H(X_{\text{Age}=Jeune}) = -\left(\frac{2}{2} \log_2 \frac{2}{2} + \frac{0}{2} \log_2 \frac{0}{2}\right) = 0$$

$$\text{Gain}(X, \text{Age}) = H(X) - \left(\frac{13}{15} H(X_{\text{Age}=Vieux}) + \frac{2}{15} H(X_{\text{Age}=Jeune})\right) \approx 0,064882 \dots$$

**Calcul du gain "Genre":**

$$\text{Gain}(X, \text{Genre}) = H(X) - \sum \frac{|X_{\text{Genre}=v}|}{|X|} H(X_{\text{Genre}=v})$$

$$H(X_{\text{Genre}=Masculin}) = -\left(\frac{7}{8} \log_2 \frac{7}{8} + \frac{1}{8} \log_2 \frac{1}{8}\right) \approx 0,54356444 \dots$$

$$H(X_{\text{Genre}=Féminin}) = -\left(\frac{4}{7} \log_2 \frac{4}{7} + \frac{3}{7} \log_2 \frac{3}{7}\right) \approx 0,98522813 \dots$$

$$\begin{aligned} \text{Gain}(X, \text{Genre}) &= H(X) - \left(\frac{8}{15} H(X_{\text{Genre}=Masculin}) + \frac{7}{15} H(X_{\text{Genre}=Féminin})\right) \\ &\approx 0,086966578 \dots \end{aligned}$$

**Calcul du gain "Hypertension":**

$$\text{Gain}(X, \text{Hypertension}) = H(X) - \sum \frac{|X_{\text{Hypertension}=v}|}{|X|} H(X_{\text{Hypertension}=v})$$

$$H(X_{\text{Hypertension}=Normale}) = -\left(\frac{5}{7} \log_2 \frac{5}{7} + \frac{2}{7} \log_2 \frac{2}{7}\right) \approx 0,86312056 \dots$$

$$H(X_{\text{Hypertension}=Grave}) = -\left(\frac{6}{8} \log_2 \frac{6}{8} + \frac{2}{8} \log_2 \frac{2}{8}\right) \approx 0,81127812 \dots$$

$$\begin{aligned} \text{Gain}(X, \text{Hypertension}) &= H(X) - \left(\frac{7}{15} H(X_{\text{Hypertension}=Normale}) + \frac{8}{15} H(X_{\text{Hypertension}=Grave})\right) \\ &\approx 0,0011694814 \dots \end{aligned}$$

**Calcul du gain "Température":**

$$\text{Gain}(X, \text{Température}) = H(X) - \sum \frac{|X_{\text{Température}=v}|}{|X|} H(X_{\text{Température}=v})$$

$$H(X_{\text{Température}=Normal}) = -\left(\frac{9}{13} \log_2 \frac{9}{13} + \frac{4}{13} \log_2 \frac{4}{13}\right) \approx 0,89049164 \dots$$

$$H(X_{\text{Température}=Grave}) = -\left(\frac{2}{2} \log_2 \frac{2}{2} + \frac{0}{2} \log_2 \frac{0}{2}\right) = 0$$

$$\begin{aligned} \text{Gain}(X, \text{Température}) &= H(X) - \left(\frac{13}{15} H(X_{\text{Temp}=Normal}) + \frac{2}{15} H(X_{\text{Temp}=Grave})\right) \\ &\approx 0,0648813186 \dots \end{aligned}$$

**Calcul du gain "ECG":**

$$\text{Gain}(X, \text{ECG}) = H(X) - \sum \frac{|X_{\text{ECG}=v}|}{|X|} H(X_{\text{ECG}=v})$$

$$H(X_{\text{ECG}=\text{Normale}}) = -\left(\frac{9}{9} \log_2 \frac{9}{9} + \frac{0}{9} \log_2 \frac{0}{9}\right) = 0$$

$$H(X_{\text{ECG}=\text{Grav}}) = -\left(\frac{2}{6} \log_2 \frac{2}{6} + \frac{4}{6} \log_2 \frac{4}{6}\right) \approx 0,91829583 \dots$$

$$\begin{aligned} \text{Gain}(X, \text{ECG}) &= H(X) - \left(\frac{9}{15} H(X_{\text{ECG}=\text{Normale}}) + \frac{6}{15} H(X_{\text{ECG}=\text{Grav}})\right) \\ &\approx 0,469322408 \dots \end{aligned}$$

**Calcul du gain "Hémoglobine":**

$$\text{Gain}(X, \text{Hémoglobine}) = H(X) - \sum \frac{|X_{\text{Hémoglobine}=v}|}{|X|} H(X_{\text{Hémoglobine}=v})$$

$$H(X_{\text{Hémoglobine}=\text{Normale}}) = -\left(\frac{11}{12} \log_2 \frac{11}{12} + \frac{1}{12} \log_2 \frac{1}{12}\right) \approx 0,41381685 \dots$$

$$H(X_{\text{Hémoglobine}=\text{Grav}}) = -\left(\frac{0}{3} \log_2 \frac{0}{3} + \frac{3}{3} \log_2 \frac{3}{3}\right) = 0$$

$$\begin{aligned} \text{Gain}(X, \text{Hémoglobine}) &= H(X) - \left(\frac{12}{15} H(X_{\text{Hémoglobine}=\text{Normale}}) + \frac{3}{15} H(X_{\text{Hémoglobine}=\text{Grav}})\right) \\ &\approx 0,50558726 \dots \end{aligned}$$

**Calcul du gain "Glucose":**

$$\text{Gain}(X, \text{Glucose}) = H(X) - \sum \frac{|X_{\text{Glucose}=v}|}{|X|} H(X_{\text{Glucose}=v})$$

$$H(X_{\text{Glucose}=\text{Normale}}) = -\left(\frac{5}{7} \log_2 \frac{5}{7} + \frac{2}{7} \log_2 \frac{2}{7}\right) \approx 0,86312056 \dots$$

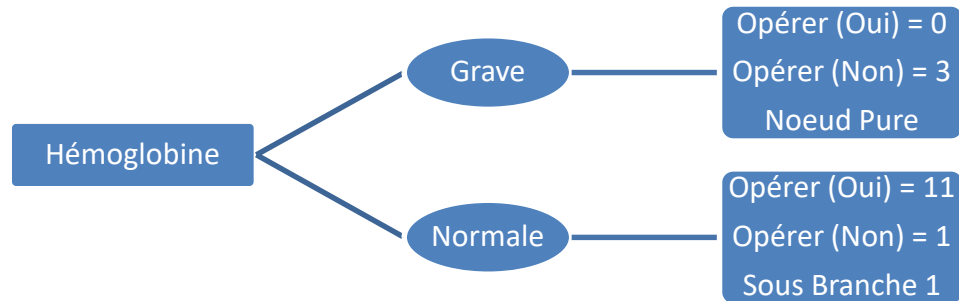
$$H(X_{\text{Glucose}=\text{Grav}}) = -\left(\frac{6}{8} \log_2 \frac{6}{8} + \frac{2}{8} \log_2 \frac{2}{8}\right) \approx 0,81127812 \dots$$

$$\begin{aligned} \text{Gain}(X, \text{Glucose}) &= H(X) - \left(\frac{7}{15} H(X_{\text{Glucose}=\text{Normale}}) + \frac{8}{15} H(X_{\text{Glucose}=\text{Grav}})\right) \\ &\approx 0,0011694813 \dots \end{aligned}$$

- Gain Age  $\approx 0,064882 \dots$
- Gain Genre  $\approx 0,086966578 \dots$
- Gain Hypertension  $\approx 0,0011694814 \dots$
- Gain Température  $\approx 0,0648813186 \dots$
- Gain ECG  $\approx 0,469322408 \dots$
- **Gain Hémoglobine  $\approx 0,50558726 \dots$**

- Gain Glucose  $\approx 0,0011694813 \dots$

Nous prenons le plus grand gain, c'est celui de l'Hémoglobine, donc il sera la racine.



Sous branche 1 :

ID	Age	Genre	Hypertension	Température	ECG	Hémoglobine	Glucose	Opérer ?
1	Vieux	Masculin	Normale	Normale	Grave	Normale	Normale	Oui
2	Jeune	Masculin	Normale	Normale	Normale	Normale	Grave	Oui
4	Vieux	Féminin	Normale	Normale	Grave	Normale	Normale	Oui
5	Vieux	Masculin	Grave	Normale	Normale	Normale	Grave	Oui
6	Vieux	Masculin	Normale	Normale	Normale	Normale	Normale	Oui
7	Vieux	Masculin	Normale	Normale	Normale	Normale	Grave	Oui
9	Vieux	Féminin	Grave	Normale	Normale	Normale	Grave	Oui
10	Vieux	Féminin	Grave	Normale	Normale	Normale	Grave	Oui
11	Vieux	Masculin	Grave	Normale	Grave	Normale	Grave	Non
12	Vieux	Masculin	Grave	Grave	Normale	Normale	Grave	Oui
14	Jeune	Féminin	Grave	Normale	Normale	Normale	Normale	Oui
15	Vieux	Masculin	Grave	Grave	Normale	Normale	Normale	Oui

Age	Genre	Hypertension	Température	ECG	Glucose	Opérer ?
Vieux = 9	Masculin = 7	Normale = 5	Normale = 9	Normale = 9	Normale = 5	Oui
Vieux = 1	Masculin = 1	Normale = 0	Normale = 1	Normale = 0	Normale = 0	Non
Jeune = 2	Féminin = 4	Grave = 6	Grave = 2	Grave = 2	Grave = 6	Oui
Jeune = 0	Féminin = 0	Grave = 1	Grave = 0	Grave = 1	Grave = 1	Non

Notre nouvelle base de données contient 11 personnes qui ont opéré, et 1 personne qui n'a pas opéré, alors l'entropie est :

$$H(X) = -\left(\frac{11}{12} \log_2 \frac{11}{12} + \frac{1}{12} \log_2 \frac{1}{12}\right) \approx 0,41381685 \dots$$

**Calcul du gain "Age":**

$$\text{Gain}(X, \text{Age}) = H(X) - \sum \frac{|X_{\text{Age}=v}|}{|X|} H(X_{\text{Age}=v})$$

$$H(X_{Age=Vieux}) = -\left(\frac{9}{10} \log_2 \frac{9}{10} + \frac{1}{10} \log_2 \frac{1}{10}\right) \approx 0,46899559 \dots$$

$$H(X_{Age=Jeune}) = -\left(\frac{2}{2} \log_2 \frac{2}{2} + \frac{0}{2} \log_2 \frac{0}{2}\right) = 0$$

$$Gain(X, Age) = H(X) - \left(\frac{10}{12} H(X_{Age=Vieux}) + \frac{2}{12} H(X_{Age=Jeune})\right) \approx 0,02298719 \dots$$

**Calcul du gain "Genre":**

$$H(X_{Genre=Masculin}) = -\left(\frac{7}{8} \log_2 \frac{7}{8} + \frac{1}{8} \log_2 \frac{1}{8}\right) \approx 0,54356444 \dots$$

$$H(X_{Genre=féminin}) = -\left(\frac{4}{4} \log_2 \frac{4}{4} + \frac{0}{4} \log_2 \frac{0}{4}\right) = 0$$

$$Gain(X, Genre) = H(X) - \left(\frac{8}{12} H(X_{Genre=Masculin}) + \frac{4}{12} H(X_{Genre=féminin})\right) \approx 0,05144055 \dots$$

**Calcul du gain "Hypertension":**

$$H(X_{Hypertension=Normale}) = -\left(\frac{5}{5} \log_2 \frac{5}{5} + \frac{0}{5} \log_2 \frac{0}{5}\right) = 0$$

$$H(X_{Hypertension=Grave}) = -\left(\frac{6}{7} \log_2 \frac{6}{7} + \frac{1}{7} \log_2 \frac{1}{7}\right) \approx 0,559167277 \dots$$

$$Gain(X, Hypertension) = H(X) - \left(\frac{5}{12} H(X_{Hypertension=Normale}) + \frac{7}{12} H(X_{Hypertension=Grave})\right) \approx 0,08763593 \dots$$

**Calcul du gain "Température":**

$$H(X_{Température=Normale}) = -\left(\frac{9}{10} \log_2 \frac{9}{10} + \frac{1}{10} \log_2 \frac{1}{10}\right) \approx 0,46899559 \dots$$

$$H(X_{Température=Grave}) = -\left(\frac{2}{2} \log_2 \frac{2}{2} + \frac{0}{2} \log_2 \frac{0}{2}\right) = 0$$

$$Gain(X, Température) = H(X) - \left(\frac{10}{12} H(X_{Température=Normale}) + \frac{2}{12} H(X_{Température=Grave})\right) \approx 0,02298719 \dots$$

**Calcul du gain "ECG":**

$$H(X_{ECG=Normal}) = -\left(\frac{9}{9} \log_2 \frac{9}{9} + \frac{0}{9} \log_2 \frac{0}{9}\right) = 0$$

$$H(X_{ECG=Grav}) = -\left(\frac{2}{3} \log_2 \frac{2}{3} + \frac{1}{3} \log_2 \frac{1}{3}\right) \approx 0,91829583 \dots$$

$$Gain(X, ECG) = H(X) - \left(\frac{7}{13} H(X_{ECG=Normale}) + \frac{6}{13} H(X_{ECG=Grav})\right) \approx 0,18424289 \dots$$

### Calcul du gain "Glucose":

$$H(X_{Glucose=Normal}) = -\left(\frac{5}{5} \log_2 \frac{5}{5} + \frac{0}{5} \log_2 \frac{0}{5}\right) = 0$$

$$H(X_{Glucose=Gr}) = -\left(\frac{6}{7} \log_2 \frac{6}{7} + \frac{1}{7} \log_2 \frac{1}{7}\right) \approx 0,59167277 \dots$$

$$Gain(X, Glucose) = H(X) - \left(\frac{5}{12} H(X_{Glucose=Normale}) + \frac{7}{12} H(X_{Glucose=Grave})\right) \\ \approx 0,06867440 \dots$$

Gain Age  $\approx 0,02298719 \dots$

Gain Genre  $\approx 0,05144055 \dots$

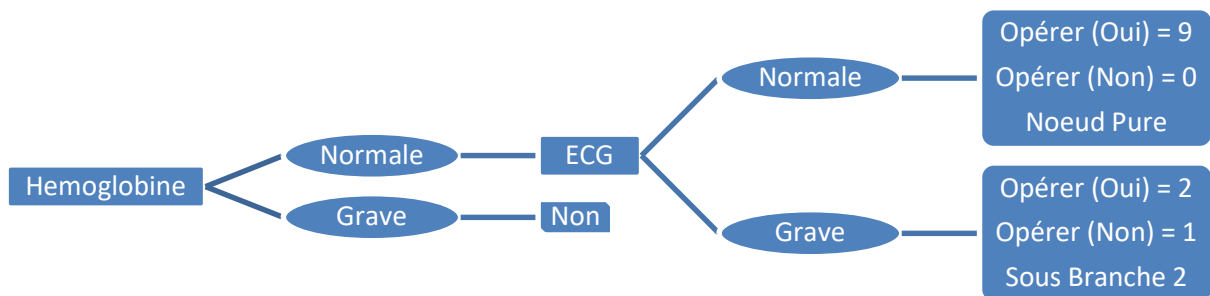
Gain Hypertension  $\approx 0,08763593 \dots$

Gain Température  $\approx 0,02298719 \dots$

**Gain ECG  $\approx 0,18424289 \dots$**

Gain Glucose  $\approx 0,06867440 \dots$

Nous prenons le plus grand gain, c'est celui de l'ECG, donc l'arbre devient :



Sous branche 2 :

ID	Age	Genre	Hypertension	Température	ECG	Hémoglobine	Glucose	Opérer ?
1	Vieux	Masculin	Normale	Normale	Grave	Normale	Normale	Oui
4	Vieux	Féminin	Normale	Normale	Grave	Normale	Normale	Oui
11	Vieux	Masculin	Grave	Normale	Grave	Normale	Grave	Non

Age	Genre	Hypertension	Température	Glucose	Opérer ?
Vieux = 2	Masculin = 1	Normale = 2	Normale = 2	Normale = 2	Oui
Vieux = 1	Masculin = 1	Normale = 0	Normale = 1	Normale = 0	Non
Jeune = 0	Féminin = 1	Grave = 0	Grave = 0	Grave = 0	Oui
Jeune = 0	Féminin = 0	Grave = 1	Grave = 0	Grave = 1	Non

Notre nouvelle base de données contient 2 personnes qui ont opéré, et 1 personne qui n'a pas opéré, alors l'entropie est :

$$H(X) = -\left(\frac{2}{3}\log_2\frac{2}{3} + \frac{1}{3}\log_2\frac{1}{3}\right) \approx 0,91829583 \dots$$

**Calcul du gain "Age":**

$$\text{Gain}(X, \text{Age}) = H(X) - \sum \frac{|X_{\text{Age}=v}|}{|X|} H(X_{\text{Age}=v})$$

$$H(X_{\text{Age}=\text{Vieux}}) = -\left(\frac{2}{3}\log_2\frac{2}{3} + \frac{1}{3}\log_2\frac{1}{3}\right) \approx 0,91829583 \dots$$

$$H(X_{\text{Age}=\text{Jeune}}) = -\left(\frac{0}{0}\log_2\frac{0}{0} + \frac{0}{0}\log_2\frac{0}{0}\right) = 0$$

$$\text{Gain}(X, \text{Age}) = H(X) - \left(\frac{3}{3} H(X_{\text{Age}=\text{Vieux}}) + \frac{0}{3} H(X_{\text{Age}=\text{Jeune}})\right) = 0$$

**Calcul du gain "Genre":**

$$H(X_{\text{Genre}=\text{Masculin}}) = -\left(\frac{1}{2}\log_2\frac{1}{2} + \frac{1}{2}\log_2\frac{1}{2}\right) = 1$$

$$H(X_{\text{Genre}=\text{Féminin}}) = -\left(\frac{1}{1}\log_2\frac{1}{1} + \frac{0}{1}\log_2\frac{0}{1}\right) = 0$$

$$\begin{aligned} \text{Gain}(X, \text{Genre}) &= H(X) - \left(\frac{2}{3} H(X_{\text{Genre}=\text{Masculin}}) + \frac{1}{3} H(X_{\text{Genre}=\text{Féminin}})\right) \\ &\approx 0,25162916 \dots \end{aligned}$$

**Calcul du gain "Hypertension":**

$$H(X_{\text{Hypertension}=\text{Normale}}) = -\left(\frac{2}{2}\log_2\frac{2}{2} + \frac{0}{0}\log_2\frac{0}{0}\right) = 0$$

$$H(X_{\text{Hypertension}=\text{Grave}}) = -\left(\frac{0}{1}\log_2\frac{0}{1} + \frac{1}{1}\log_2\frac{1}{1}\right) = 0$$

$$\begin{aligned} \text{Gain}(X, \text{Hypertension}) &= H(X) - \left(\frac{2}{3} H(X_{\text{Hypertension}=\text{Normale}}) + \frac{1}{3} H(X_{\text{Hypertension}=\text{Grav}})\right) \\ &\approx 0,91829583 \dots \end{aligned}$$

**Calcul du gain "Température":**

$$H(X_{\text{Température}=\text{Normale}}) = -\left(\frac{2}{3}\log_2\frac{2}{3} + \frac{1}{3}\log_2\frac{1}{3}\right) \approx 0,91829583 \dots$$

$$H(X_{\text{Température}=\text{Grav}}) = -\left(\frac{0}{0}\log_2\frac{0}{0} + \frac{0}{0}\log_2\frac{0}{0}\right) = 0$$

$$\begin{aligned} \text{Gain}(X, \text{Température}) &= H(X) - \left(\frac{3}{3} H(X_{\text{Température}=\text{Norm}}) + \frac{0}{3} H(X_{\text{Température}=\text{Grav}})\right) = 0 \end{aligned}$$

**Calcul du gain "Glucose":**

$$H(X_{\text{Glucose}=\text{Normal}}) = -\left(\frac{2}{2}\log_2\frac{2}{2} + \frac{0}{0}\log_2\frac{0}{0}\right) = 0$$

$$H(X_{Glucose=Grav}) = -\left(\frac{0}{1} \log_2 \frac{0}{1} + \frac{1}{1} \log_2 \frac{1}{1}\right) = 0$$

$$Gain(X, Glucose) = H(X) - \left(\frac{2}{3} H(X_{Glucose=Normale}) + \frac{1}{3} H(X_{Glucose=Grav})\right) \\ \approx 0,91829583$$

Gain Age = 0

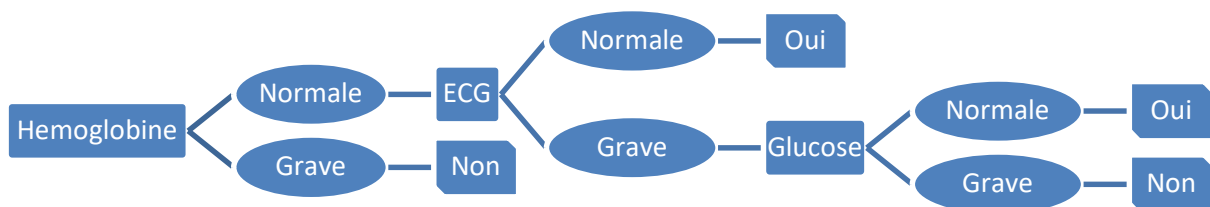
Gain Genre  $\approx 0,25162916$  ...

Gain Hypertension  $\approx 0,91829583$  ...

Gain Température = 0

Gain Glucose  $\approx 0,91829583$  ...

Nous prenons le plus grand gain, c'est celui de l'hypertension et du Glucose, nous choisissons l'un des deux, par exemple celui du Glucose, donc l'arbre devient :



Arrêt de l'algorithme ID3, Il n'y a plus de caractéristiques pour diviser les données.

## B. Solution avec l'algorithme CART

L'algorithme commence par évaluer chaque attribut disponible pour déterminer lequel serait le meilleur choix pour diviser les données. Pour ce faire, nous examinons les différentes valeurs de chaque attribut par rapport aux deux classes cibles, c'est-à-dire "Opérer" ou "Ne pas opérer". Le tableau suivant présente le nombre de cas "Oui" et "Non" pour chaque combinaison de valeurs d'attributs :

Age	Genre	Hypertension	Température	ECG	Hémoglobine	Glucose	Opérer ?
Vieux = 9	Masculin = 7	Normale = 5	Normale = 9	Normale = 9	Normale = 11	Normale = 5	Oui
Vieux = 4	Masculin = 1	Normale = 2	Normale = 4	Normale = 0	Normale = 1	Normale = 2	Non
Jeune = 2	Féminin = 4	Grave = 6	Grave = 2	Grave = 2	Grave = 0	Grave = 6	Oui
Jeune = 0	Féminin = 3	Grave = 2	Grave = 0	Grave = 4	Grave = 3	Grave = 2	Non

$$Gini - index (Attribut = Valeur) = 1 - \sum (P_i)^2$$

$$Gini - index (Attribut) = \sum P_v * GI(v)$$



Calcule de l'indice de Gini pour l'attribut « Age » :

$$Gini - index (Age = Vieux) = 1 - \left( \left( \frac{9}{13} \right)^2 + \left( \frac{4}{13} \right)^2 \right) \approx 0,42603550 \dots$$

$$Gini - index (Age = jeune) = 1 - \left( \left( \frac{2}{2} \right)^2 + \left( \frac{0}{2} \right)^2 \right) = 0$$

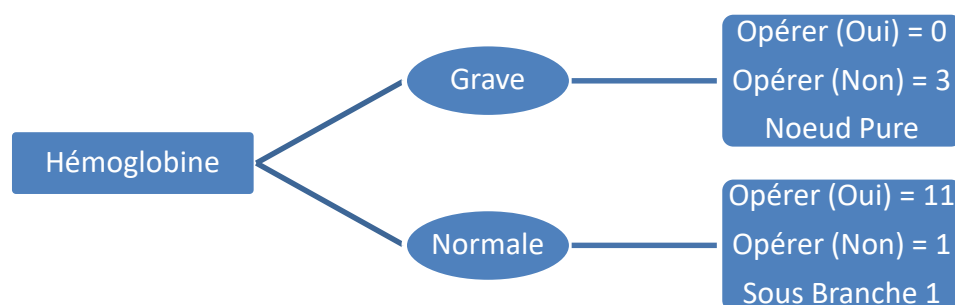
$$Gini - index (Age) = \left( 0,4260355 * \frac{13}{15} \right) + \left( 0 * \frac{2}{15} \right) \approx 0,36923076 \dots$$

Attribut	Indice de Gini	Valeur
Age	$[(1 - ((9/13)^2 + (4/13)^2)) * 13/15] + [(1 - ((2/2)^2 + (0/2)^2)) * 2/15]$	0,369
Genre	$[(1 - ((7/8)^2 + (1/8)^2)) * 8/15] + [(1 - ((4/7)^2 + (3/7)^2)) * 7/15]$	0,345
Hypertension	$[(1 - ((5/7)^2 + (2/7)^2)) * 7/15] + [(1 - ((6/8)^2 + (2/8)^2)) * 8/15]$	0,39
Température	$[(1 - ((9/13)^2 + (4/13)^2)) * 13/15] + [(1 - ((2/2)^2 + (0/2)^2)) * 2/15]$	0,369
ECG	$[(1 - ((9/9)^2 + (0/9)^2)) * 9/15] + [(1 - ((2/6)^2 + (4/6)^2)) * 6/15]$	0,177
Hémoglobine	$[(1 - ((11/12)^2 + (1/12)^2)) * 12/15] + [(1 - ((0/3)^2 + (3/3)^2)) * 3/15]$	<b>0,122</b>
Glucose	$[(1 - ((5/7)^2 + (2/7)^2)) * 7/15] + [(1 - ((6/8)^2 + (2/8)^2)) * 8/15]$	0,39

L'attribut qui minimise l'indice de Gini est sélectionné comme variable de partitionnement, dans notre cas c'est l'hémoglobine.

Une fois l'attribut de partitionnement sélectionné, l'algorithme divise les données en sous-ensembles en fonction des différentes valeurs de cet attribut.

Chaque sous-ensemble représente un nœud ou une branche de l'arbre.



ID	Age	Genre	Hypertension	Température	ECG	Hémoglobine	Glucose	Opérer ?
1	Vieux	Masculin	Normale	Normale	Grave	Normale	Normale	Oui
2	Jeune	Masculin	Normale	Normale	Normale	Normale	Grave	Oui
4	Vieux	Féminin	Normale	Normale	Grave	Normale	Normale	Oui
5	Vieux	Masculin	Grave	Normale	Normale	Normale	Grave	Oui
6	Vieux	Masculin	Normale	Normale	Normale	Normale	Normale	Oui
7	Vieux	Masculin	Normale	Normale	Normale	Normale	Grave	Oui
9	Vieux	Féminin	Grave	Normale	Normale	Normale	Grave	Oui

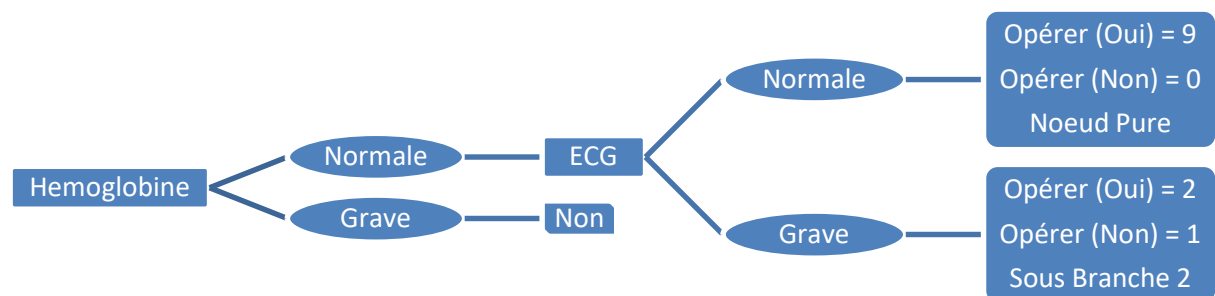
10	Vieux	Féminin	Grave	Normale	Normale	Normale	Grave	Oui
11	Vieux	Masculin	Grave	Normale	Grave	Normale	Grave	Non
12	Vieux	Masculin	Grave	Grave	Normale	Normale	Grave	Oui
14	Jeune	Féminin	Grave	Normale	Normale	Normale	Normale	Oui
15	Vieux	Masculin	Grave	Grave	Normale	Normale	Normale	Oui

Age	Genre	Hypertension	Température	ECG	Glucose	Opérer ?
Vieux = 9	Masculin = 7	Normale = 5	Normale = 9	Normale = 9	Normale = 5	Oui
Vieux = 1	Masculin = 1	Normale = 0	Normale = 1	Normale = 0	Normale = 0	Non
Jeune = 2	Féminin = 4	Grave = 6	Grave = 2	Grave = 2	Grave = 6	Oui
Jeune = 0	Féminin = 0	Grave = 1	Grave = 0	Grave = 1	Grave = 1	Non

Calcule de l'indice de Gini

Attribut	Indice de Gini	Valeur
Age	$[(1 - ((9/10)^2 + (1/10)^2)) * 10/12] + [(1 - ((2/2)^2 + (0/2)^2)) * 2/12]$	0,15
Genre	$[(1 - ((7/8)^2 + (1/8)^2)) * 8/12] + [(1 - ((4/4)^2 + (0/4)^2)) * 4/12]$	0,1458
Hypertension	$[(1 - ((5/5)^2 + (0/5)^2)) * 5/12] + [(1 - ((6/7)^2 + (1/7)^2)) * 7/12]$	0,1428
Température	$[(1 - ((9/10)^2 + (1/10)^2)) * 10/12] + [(1 - ((2/2)^2 + (0/2)^2)) * 2/12]$	0,15
ECG	$[(1 - ((9/9)^2 + (0/9)^2)) * 9/12] + [(1 - ((2/3)^2 + (1/3)^2)) * 3/12]$	<b>0,1111</b>
Glucose	$[(1 - ((5/5)^2 + (0/5)^2)) * 5/12] + [(1 - ((6/7)^2 + (1/7)^2)) * 7/12]$	0,1428

Nous remarquons que l'attribut qui minimise l'indice de Gini c'est « ECG », donc l'arbre devient :



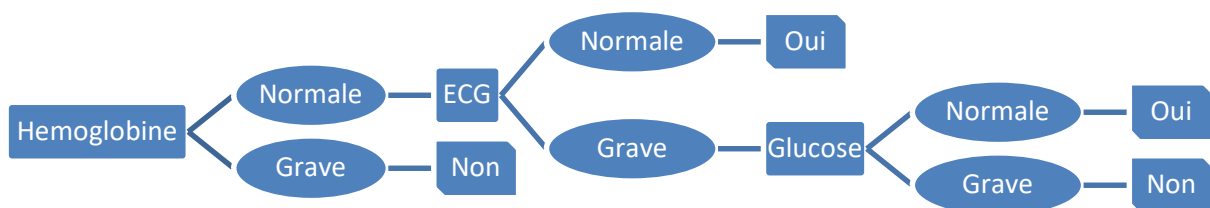
ID	Age	Genre	Hypertension	Température	ECG	Hémoglobine	Glucose	Opérer ?
1	Vieux	Masculin	Normale	Normale	Grave	Normale	Normale	Oui
4	Vieux	Féminin	Normale	Normale	Grave	Normale	Normale	Oui
11	Vieux	Masculin	Grave	Normale	Grave	Normale	Grave	Non

Age	Genre	Hypertension	Température	Glucose	Opérer ?
Vieux = 2	Masculin = 1	Normale = 2	Normale = 2	Normale = 2	Oui
Vieux = 1	Masculin = 1	Normale = 0	Normale = 1	Normale = 0	Non
Jeune = 0	Féminin = 1	Grave = 0	Grave = 0	Grave = 0	Oui
Jeune = 0	Féminin = 0	Grave = 1	Grave = 0	Grave = 1	Non

## Calcul de l'indice de Gini

Attribut	Indice de Gini	Valeur
Age	$[(1 - ((2/3)^2 + (1/3)^2)) * 3/3] + [(1 - ((0/0)^2 + (0/0)^2)) * 0/3]$	Non défini
Genre	$[(1 - ((1/2)^2 + (1/2)^2)) * 2/3] + [(1 - ((1/1)^2 + (0/1)^2)) * 1/3]$	0,33
Hypertension	$[(1 - ((2/2)^2 + (0/2)^2)) * 2/3] + [(1 - ((0/1)^2 + (1/1)^2)) * 1/3]$	<b>0</b>
Température	$[(1 - ((2/3)^2 + (1/3)^2)) * 3/3] + [(1 - ((0/0)^2 + (0/0)^2)) * 0/3]$	Non défini
Glucose	$[(1 - ((2/2)^2 + (0/2)^2)) * 2/3] + [(1 - ((0/1)^2 + (1/1)^2)) * 1/3]$	<b>0</b>

Nous remarquons qu'il y'a deux attributs qui minimisent l'indice de Gini, nous choisissons l'un des deux, par exemple le Glucose, donc l'arbre devient :



Arrêt de l'algorithme CART, Il n'y a plus de caractéristiques pour diviser les données.

Utilisation des arbres de décision (ID3 et CART) pour prédire si un nouveau patient avec les caractéristiques suivantes doit subir une opération : âge = "Vieux", genre = "Féminin", hypertension = "Grave", température = "Grave", ECG = "Normale", hémoglobine = "Normale", glucose = "Normale", en suivant le chemin approprié dans l'arbre pour prendre notre décision.

