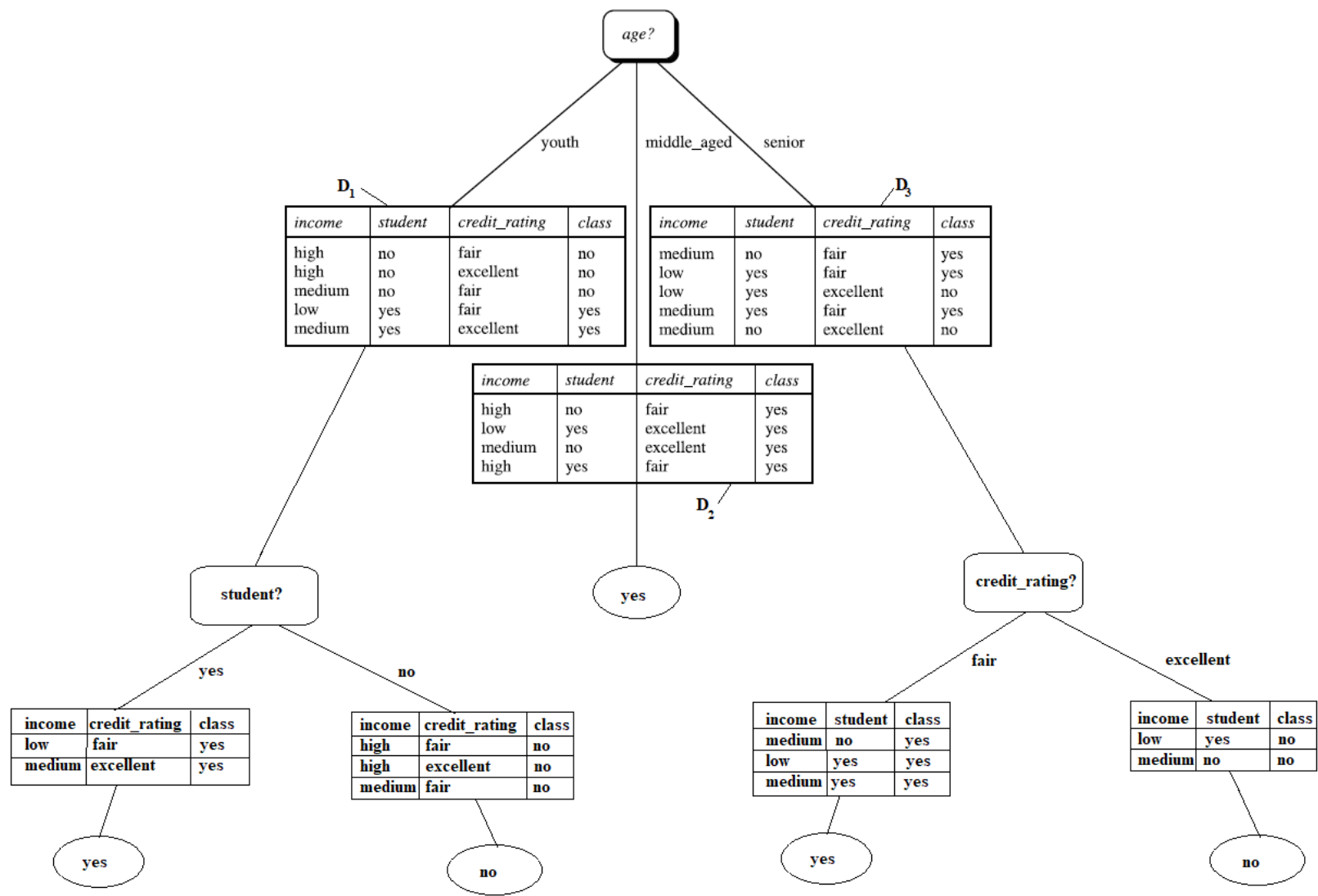


Name: Lester Hernandez Alfonso

Complete Decision Tree:



## Process and Calculations:

Partition  $D_2$ :

1

\* All tuples from same class. Add leaf node 'yes' to path of  $D_2$ .

Partition  $D_1$ :

Information Gain Calculation:

$$\text{Info}(D_1) = -\frac{2}{5} \log_2\left(\frac{2}{5}\right) - \frac{3}{5} \log_2\left(\frac{3}{5}\right)$$

$$\text{Info}(D_1) \approx \underline{0.971 \text{ bits}}$$

$$\begin{aligned} \text{Info}_{\text{income}}(D_1) &= \frac{1}{5} \times \left(-\frac{1}{1} \log_2\left(\frac{1}{1}\right)\right) \\ &\quad + \frac{2}{5} \times \left(-\frac{1}{2} \log_2\left(\frac{1}{2}\right) - \frac{1}{2} \log_2\left(\frac{1}{2}\right)\right) \\ &\quad + \frac{2}{5} \times \left(-\frac{2}{2} \log_2\left(\frac{2}{2}\right)\right) \\ &= \underline{0.4 \text{ bits}} \end{aligned}$$

$$\begin{aligned} \text{Info}_{\text{student}}(D_1) &= \frac{2}{5} \times \left(-\frac{2}{2} \log_2\left(\frac{2}{2}\right)\right) \\ &\quad + \frac{3}{5} \times \left(-\frac{3}{3} \log_2\left(\frac{3}{3}\right)\right) \\ &= \underline{0 \text{ bits}} \end{aligned}$$

$$\begin{aligned} \text{Info}_{\text{credit\_rating}}(D_1) &= \frac{3}{5} \times \left(-\frac{1}{3} \log_2\left(\frac{1}{3}\right) - \frac{2}{3} \log_2\left(\frac{2}{3}\right)\right) \\ &\quad + \frac{2}{5} \times \left(-\frac{1}{2} \log_2\left(\frac{1}{2}\right) - \frac{1}{2} \log_2\left(\frac{1}{2}\right)\right) \\ &\approx \underline{0.951 \text{ bits}} \end{aligned}$$

$$\begin{aligned} \text{GAIN}(\text{income}) &= \text{Info}(D_1) - \text{Info}_{\text{income}}(D_1) \\ &= 0.971 \text{ bits} - 0.4 \text{ bits} \\ &= \underline{0.571} \end{aligned}$$

$$\begin{aligned} \text{GAIN}(\text{student}) &= \text{Info}(D_1) - \text{Info}_{\text{student}}(D_1) \\ &= 0.971 \text{ bits} - 0 \text{ bits} \\ &= \underline{0.971 \text{ bits}} \end{aligned}$$

$$\begin{aligned} \text{GAIN}(\text{credit\_rating}) &= \text{Info}(D_1) - \text{Info}_{\text{credit\_rating}}(D_1) \\ &= 0.971 \text{ bits} - 0.951 \text{ bits} \\ &= \underline{0.020} \end{aligned}$$

\*'Student' is selected as the splitting Attribute because it has the highest information gain.

- We create a new node labeled 'student?' on the path of  $D_1$ .
- Branches are grown from the new node with values 'yes' and 'no' for possible outcomes.
- The partition for 'yes' is shown below:

income	credit-rating	class
low	fair	yes
medium	excellent	yes

- The partition for 'no' is shown below:

income	credit-rating	class
high	fair	no
high	excellent	no
medium	fair	no

\* Since both partitions contain only tuples from the same class, we add leaf nodes 'yes' and 'no' respectively.



Partition  $D_3$ :

Information Gain Calculation:

$$\begin{aligned}\text{Info}(D_3) &= -\frac{3}{5} \log_2\left(\frac{3}{5}\right) - \frac{2}{5} \log_2\left(\frac{2}{5}\right) \\ &= \underline{0.971}\end{aligned}$$

$$\begin{aligned}\text{Info}_{\text{income}}(D_3) &= \frac{2}{5} \times \left(-\frac{1}{2} \log_2\left(\frac{1}{2}\right) - \frac{1}{2} \log_2\left(\frac{1}{2}\right)\right) \\ &\quad + \frac{3}{5} \times \left(-\frac{2}{3} \log_2\left(\frac{2}{3}\right) - \frac{1}{3} \log_2\left(\frac{1}{3}\right)\right) \\ &\approx \underline{0.951 \text{ bits}}\end{aligned}$$

$$\begin{aligned}\text{Info}_{\text{student}}(D_3) &= \frac{3}{5} \times \left(-\frac{2}{3} \log_2\left(\frac{2}{3}\right) - \frac{1}{3} \log_2\left(\frac{1}{3}\right)\right) \\ &\quad + \frac{2}{5} \times \left(-\frac{1}{2} \log_2\left(\frac{1}{2}\right) - \frac{1}{2} \log_2\left(\frac{1}{2}\right)\right) \\ &\approx \underline{0.951 \text{ bits}}\end{aligned}$$

$$\begin{aligned}\text{Info}_{\text{credit\_rating}}(D_3) &= \frac{3}{5} \left(-\frac{3}{3} \log_2\left(\frac{3}{3}\right)\right) \\ &\quad + \frac{2}{5} \left(-\frac{2}{2} \log_2\left(\frac{2}{2}\right)\right) \\ &= \underline{0 \text{ bits}}\end{aligned}$$

$$\begin{aligned}\text{GAIN}(\text{income}) &= \text{Info}(D_3) - \text{Info}_{\text{income}}(D_3) \\ &= 0.971 \text{ bits} - 0.951 \text{ bits} = \underline{0.020 \text{ bits}}\end{aligned}$$

$$\begin{aligned}\text{GAIN}(\text{student}) &= \text{Info}(D_3) - \text{Info}_{\text{student}}(D_3) \\ &= 0.971 \text{ bits} - 0.951 \text{ bits} = \underline{0.020 \text{ bits}}\end{aligned}$$

$$\begin{aligned}\text{GAIN}(\text{credit\_rating}) &= \text{Info}(D_3) - \text{Info}_{\text{credit\_rating}}(D_3) \\ &= 0.971 \text{ bits} - 0 \text{ bits} \\ &= \underline{0.971 \text{ bits}}\end{aligned}$$

\* 'Credit-Rating' is selected as the splitting <sup>4</sup> Attribute because it has the highest information gain.

- We create a new node labeled 'credit-rating?' on the path of  $D_3$ .
- Branches are grown from the new node with values 'fair' and 'excellent' for possible outcomes.
- The partition for 'fair' is shown below:

income	student	class
medium	no	yes
low	yes	yes
medium	yes	yes

- The partition for 'excellent' is shown below:

income	student	class
low	yes	no
medium	no	no

\* Since both partitions contain only tuples from the same class, we add leaf nodes 'yes' and 'no' along the respective path of the corresponding partition.

\* All leaf nodes have been generated. The decision tree is complete.