Krista Miller

Data Mining Exercise 6

Part 1:

	INCOME	STUDENT	CREDIT	CLASS
YOUTH	HIGH	NO	FAIR	NO
YOUTH	HIGH	NO	EXCELLENT	NO
YOUTH	MED	NO	FAIR	NO
YOUTH	LOW	YES	FAIR	YES
YOUTH	MED	YES	EXCELLENT	YES

GAIN

$$INFO(D) = -\frac{3}{5}\log_2\left(\frac{3}{5}\right) - \frac{2}{5}\log_2\left(\frac{2}{5}\right) = 0.971$$

$$INFO_{(student)}(D) = \underbrace{\frac{3}{5}\left(\frac{-3}{3}\log_2\frac{3}{3}\right)}_{NO} + \underbrace{\frac{2}{5}\left(\frac{-2}{2}\log_2\left(\frac{2}{2}\right)\right)}_{YES} = 0+0$$

$$INFO_{(credit)}(D) = \underbrace{\frac{3}{5}\left(\frac{-2}{3}\log_2\frac{2}{3} - \frac{1}{3}\log_2\frac{1}{3}\right)}_{FAIR} + \underbrace{\frac{2}{5}\left(\frac{-1}{2}\log_2\frac{1}{2} - \frac{1}{2}\log_2\frac{1}{2}\right)}_{EXCELLENT} = .551 + 0.4 = 0.951$$

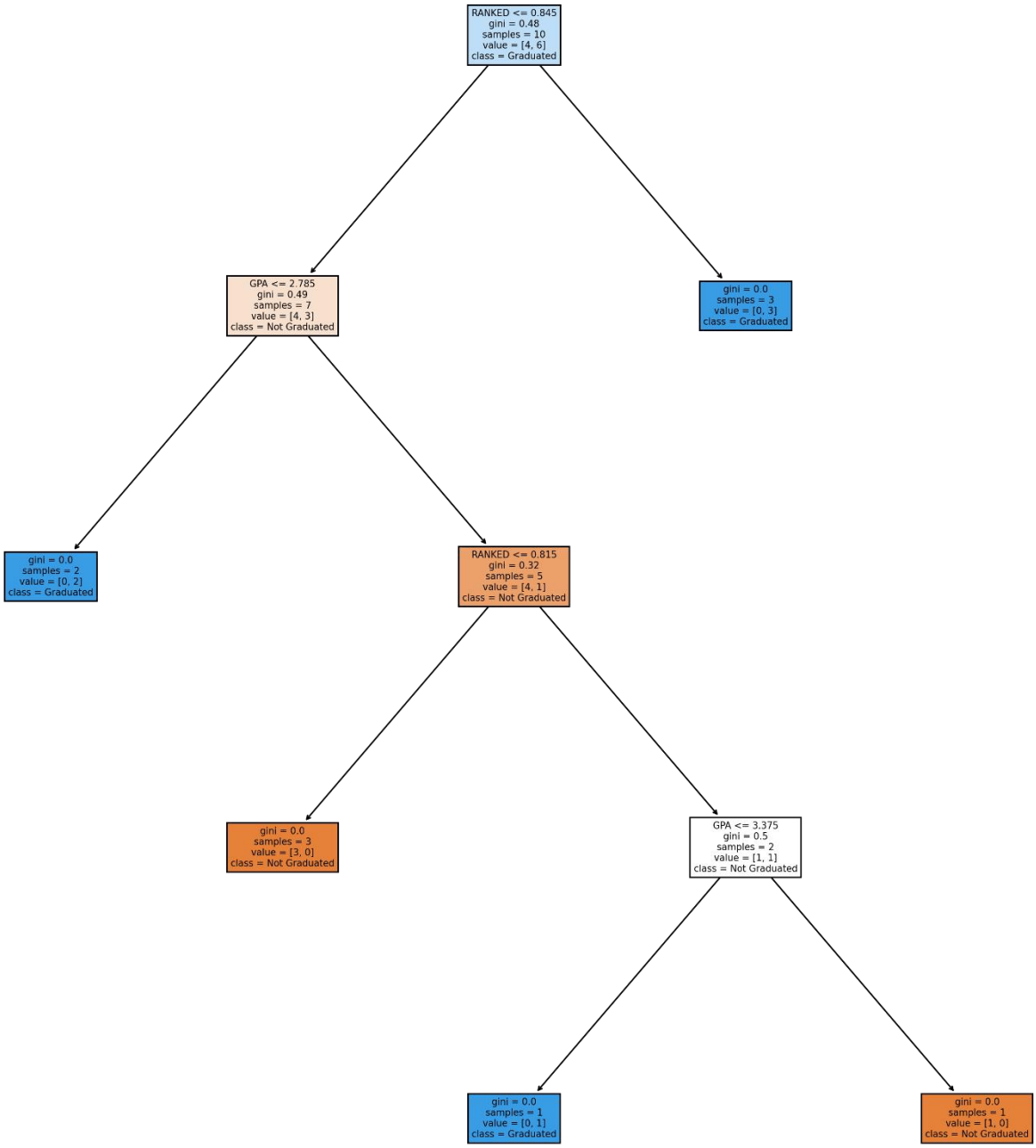
$$INFO_{(income)}(D) = \underbrace{\frac{2}{5}\left(\frac{-2}{2}\log_2\frac{2}{2}\right)}_{HIGH} + \underbrace{\frac{1}{5}\left(-1\log_2 1\right)}_{LOW} + \underbrace{\frac{2}{5}\left(\underbrace{\frac{-1}{2}\log_2\frac{1}{2}}_{NO} - \underbrace{\frac{1}{2}\log_2\frac{1}{2}}_{YES}\right)}_{MED} = 0+0+0.4 = 0.4$$

$$GAIN_{(income)} = INFO(D) - INFO_{(income)}(D) = 0.971 - 0.4 = 0.571$$

$$GAIN_{(student)} = INFO(D) - INFO_{(student)}(D) = 0.971 - 0 = 0.971$$

$$GAIN_{(credit)} = INFO(D) - INFO_{(credit)}(D) = 0.971 - 0.951 = 0.020$$

Part 2:



Root + decision node RANK >= 0.845: True= 3 students graduated (leaf/terminal node); False= 7 students

Decision node GPA <= 2.785: True= 2 students graduated (leaf/terminal node); False= 5 students

Decision node RANK <= 0.815: True= 3 students not graduated(leaf/terminal node); False= 2 students

Decision node GPA <= 3.375: True= 1 student graduated(leaf/terminal node); False= 1 student(leaf/terminal node)