July 11, 2018

Course: CIS570 – Business Intelligence

Name: Robert Palumbo

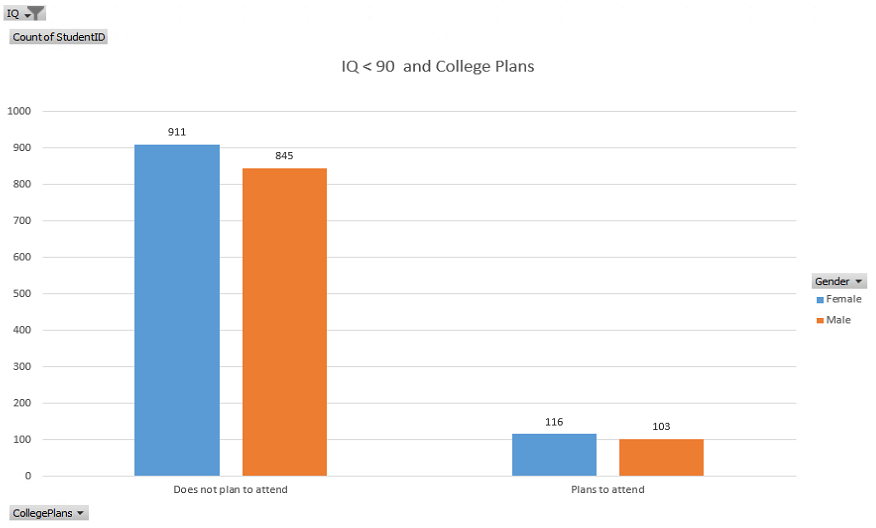
Assignment: Final Exam

Due Date: Wednesday, July 11@ 11:59pm

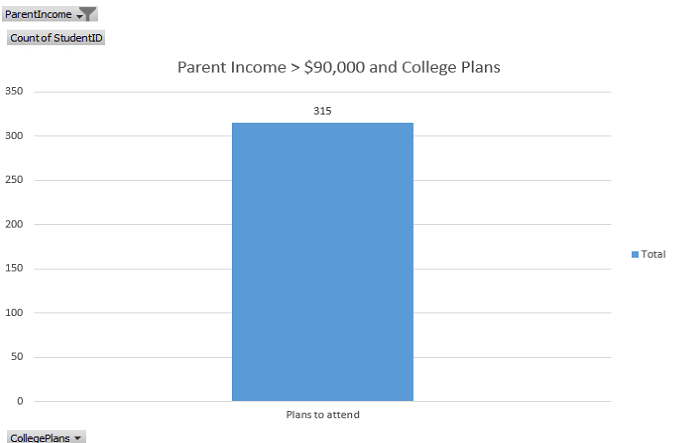
Re: Answers to Final Exam questions.

* 1. How many students are in the sample data set? 8000 Students
  2. What percentage of them have plans to attend college? (2596 / 8000) = 32.45%
  3. What percentage do not have plans to attend college? (5404 / 8000) = 67.55%
  4. How many male students are in the sample? 3874
  5. What percentage of the female students have plans to attend college? (1184 / 8000) = 14.8%
  6. What is the mean (average) IQ of the students in the sample dataset? 99.57
  7. If a student’s IQ is less than 90, is he/she more or less likely to have plans to attend college?

Student is **LESS** likely to have plans to attend college.



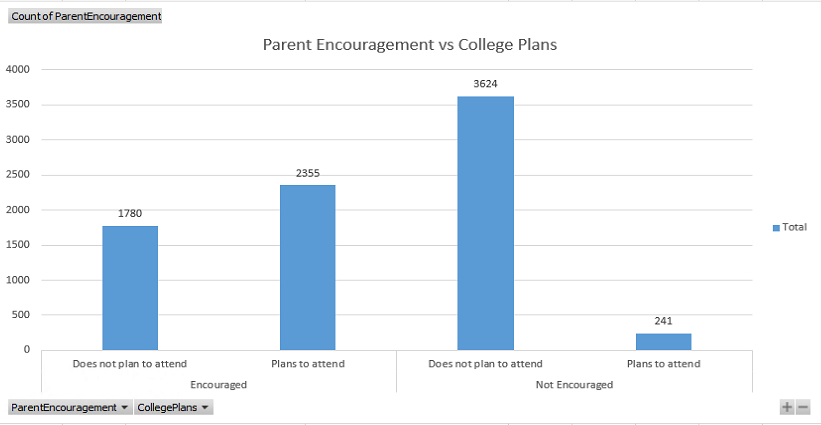
* 1. What is the minimum parental income in the sample dataset? $4500.00
  2. If a student’s parents have income greater than $75,000, is he/she more or less likely to have plans to attend college? Offer support for your answer. (2 points)



* 1. Is Parental Encouragement likely to be a good predictor of a student’s plans to attend college or not? Explain.

With reference to the following figure, in my opinion Parental encouragement is only a good predictor of a student’s plans to NOT attend college. In this case 45.3% of students plan on NOT attending while 3.01% plan on attending when a parent does NOT encourage attendance. There is a decidedly heavy bias towards NOT attending college when a parent does not encourage it.

However, when parents do encourage students to attend college, 29.44% will plan ON attending whereas 22.25% will plan on NOT attending. This is a fairly close margin in this case. In my opinion here, a student is sort of on the fence as to whether they will plan on attending college or not if a parent encourages it, thus for this case parental encouragement does not make a good predictor.



Use the data to build a) Decision-Tree and b) Neural Network models to predict whether a student has plans to attend college or not. Use gender, IQ, parental income and encouragement as inputs. Reserve 25% of the data for testing and set the “HoldoutSeed” property to 75.

* 1. Based on the results of the Neural Network model, which input variables (e.g., IQ) and their associated values (e.g., between 85 and 100) have the strongest influence on a female student’s plans to a) attend college and b) not attend college. What are the probabilities for the two outcomes? (3 points)

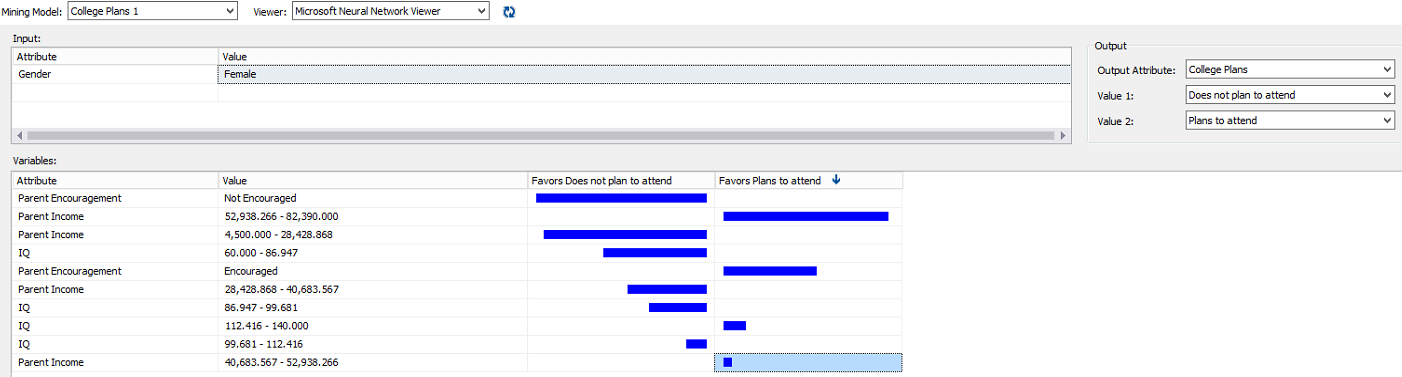
Refer to the figure below for verification. The input variables with the strongest influence for females to **ATTEND** college are:

* + 1. Parental Income
       1. Range: **52,812.45 – 82,390**
       2. Percentages
          1. Plans to attend: **60.28%**
          2. Does not plan to attend: **39.69%**
    2. Parent Encouragement
       1. Value: **Encouraged**
       2. Percentages
          1. Plans to attend: **41.19%**
          2. Does not plan to attend: **58.76%**
    3. IQ
       1. Range: **112.449 – 140.0**
       2. Percentages
          1. Plans to attend: **24.27%**
          2. Does not plan to attend: **75.71%**
    4. Parent Income
       1. Range: **40,683.56 – 52,938.26**
       2. Percentages
          1. Plans to attend: **21.58%**
          2. Does not plan to attend: **78.39%**

The input variables with the strongest influence for females to **NOT ATTEND** college are:

* + 1. Parent Encouragement
       1. Value: **NOT Encouraged**
       2. Percentages
          1. Plans to attend: **3.76%**
          2. Does not plan to attend: **96.21%**
    2. Parental Income
       1. Range: **4,500 – 28,428.86**
       2. Percentages
          1. Plans to attend: **4.04%**
          2. Does not plan to attend: **95.94%**
    3. IQ
       1. Range: **60.0 – 86.947**
       2. Percentages
          1. Plans to attend: **7.51%**
          2. Does not plan to attend: **92.47%**
    4. Parental Income
       1. Range: **28,428.86 – 40,683.56**
       2. Percentages
          1. Plans to attend: **9.49%**
          2. Does not plan to attend: **90.49%**
    5. IQ
       1. Range: **86.947 – 99.681**
       2. Percentages
          1. Plans to attend: **11.79%**
          2. Does not plan to attend: **88.18%**
    6. IQ
       1. Range: **99.681 – 112.416**
       2. Percentages
          1. Plans to attend: **16.61%**
          2. Does not plan to attend: **83.37%**

Support picture.



* 1. Based on the results of the Neural Network model, which input variables (e.g., IQ) and their associated values (e.g., between 85 and 100) have the strongest influence on a male student’s plans to a) attend college and not attend college. What are the probabilities for the two outcomes? (3 points)

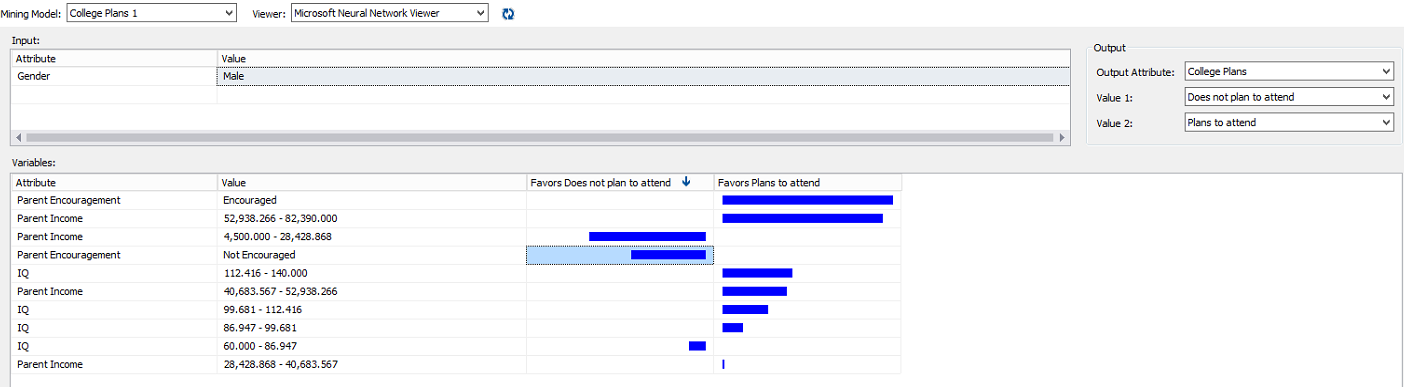
Refer to the figure below for verification. The input variables with the strongest influence for males to **ATTEND** college are:

* + 1. Parent Encouragement
       1. Value: **Encouraged**
       2. Percentages
          1. Plans to attend: **51.75%**
          2. Does not plan to attend: **48.23%**
    2. Parental Income
       1. Range: **52,938.26 – 82,390**
       2. Percentages
          1. Plans to attend: **49.63%**
          2. Does not plan to attend: **50.34%**
    3. IQ
       1. Range: **112.416 – 140.0**
       2. Percentages
          1. Plans to attend: **31.41%**
          2. Does not plan to attend: **68.57%**
    4. Parental Income
       1. Range: **40,683.56 – 52,938.26**
       2. Percentages
          1. Plans to attend: **30.39%**
          2. Does not plan to attend: **69.59%**
    5. IQ
       1. Range: **99.681– 112.416**
       2. Percentages
          1. Plans to attend: **27.09%**
          2. Does not plan to attend: **72.89%**
    6. IQ
       1. Range: **86.947 – 99.681**
       2. Percentages
          1. Plans to attend: **23.15%**
          2. Does not plan to attend: **76.83%**
    7. Parental Income
       1. Range: **38,428.86 – 40,683.56**
       2. Percentages
          1. Plans to attend: **20.44%**
          2. Does not plan to attend: **79.54%**

The input variables with the strongest influence for males to **NOT ATTEND** college are:

* + 1. Parental Income
       1. Range: **4,500 – 28,428.86**
       2. Percentages
          1. Plans to attend: **8.5%**
          2. Does not plan to attend: **91.47%**
    2. Parent Encouragement
       1. Value: **NOT Encouraged**
       2. Percentages
          1. Plans to attend: **11.69%**
          2. Does not plan to attend: **88.29%**
    3. IQ
       1. Range: **60.0 – 86.947**
       2. Percentages
          1. Plans to attend: **17.82%**
          2. Does not plan to attend: **82.15%**

Support picture.



* 1. Based on the results of the Decision Tree model, identify the node (with a minimum of 100 cases) that has students with the highest probability of not planning to attend college? What is the probability? What is the rule for this node? (3 points)

Node: **Parent Income < 43445**

Probability: **97.18%**

Node Rule: **Parent Encouragement = 'Not Encouraged' and IQ < 100 and Parent Income < 43445**

* 1. Based on the results of the Decision Tree model, identify the node that has students with the highest probability of planning to attend college? What is the probability? What is the rule for this node? (3 points)

Node: **Parent Income >= 67591**

Probability: **99.97%**

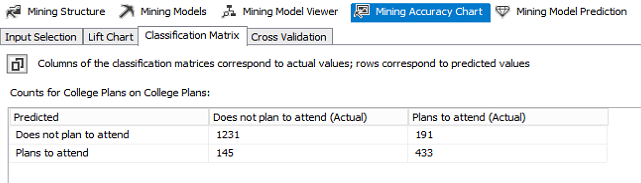
Node Rule: **Parent Encouragement = 'Encouraged' and Parent Income >= 67591 and < 74601**

* 1. Based on the results of the Decision Tree model, which inputs are the weakest and strongest predictors of “College Plans”? (1 point)

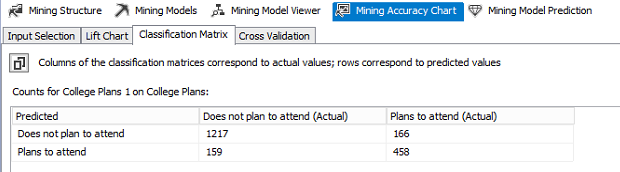
Weakest Predictor: **IQ**

Strongest Predictor: **Parent Encouragement**

* 1. Complete the following table for the a) Decision Tree and b) Neural Network models. (2 points)
     1. Decision Tree



* + 1. Neural Network



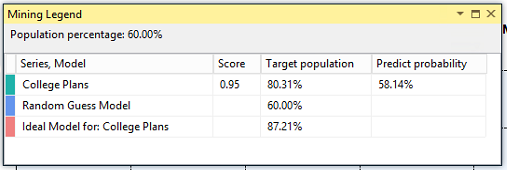
* 1. Based on the values in the classification tables, which model has a lower percentage of cases predicted incorrectly? What is the percentage? (3 points)

Decision Tree: (191 + 145) / 2000 = 16.8%

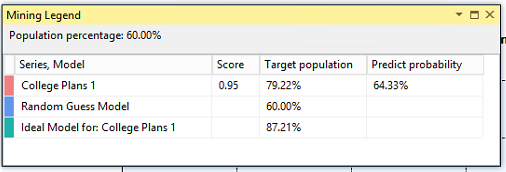
Neural Network: (166 + 159) / 2000 = **16.25**%

**The Neural Network with 16.25% predicted incorrectly is the lowest.**

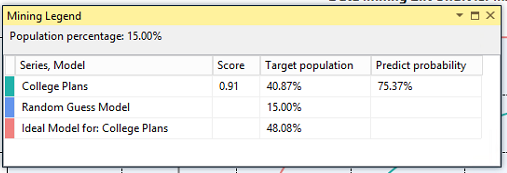
* 1. For the Predict Value, “Does not plan to attend”, at 60% of the overall population, what percentage of the target population is correctly predicted by a) the ideal model, b) the Decision Tree model, and c) the Neural Network model? (3 points)
     1. Decision Tree



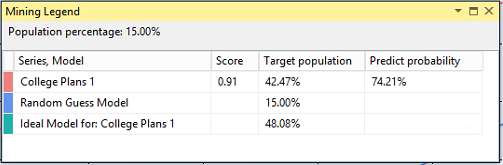
* + 1. Neural Network



* 1. For the Predict Value, “Plans to attend”, at 15% of the overall population, what percentage of the target population is correctly predicted by a) the ideal model, b) the Decision Tree model, and c) the Neural Network model? (3 points)
     1. Decision Tree



* + 1. Neural Network



* 1. Ram U will pay you $10 for each mailed brochure that reaches a high school graduate who “Plans to attend college”. Assume that the population of high school graduates is 12,000, the fixed cost for mailing the brochures is $2,000 and it costs $2 to mail each brochure. How many students in this population will have plans to attend college? If your objective is to maximize profit, a) which model’s (i.e., Decision Tree or Neural Network) recommendation will you follow, b) to how many students should you mail the brochures, and c) what will be your profit? (4 points)
     1. Decision Tree is the model that will maximize profit: 41% of population, profit = **$20,260.00**
        1. Of the original test data set of 8000 students there are 2,596 planning on attending college resulting in a sample percentage of: (2596 / 8000) = 32.45%

From the scoring population of 12,000, we can expect there to be (0.3245 \* 12000) = 3,894 students planning on attending college.

From the lift chart at 41%for the decision tree model we get at target population % of: 85.74%

**If we follow the model we should mail the brochure to:**

**0.8574 \* 3,894 = 3,338 students (rounded down)**

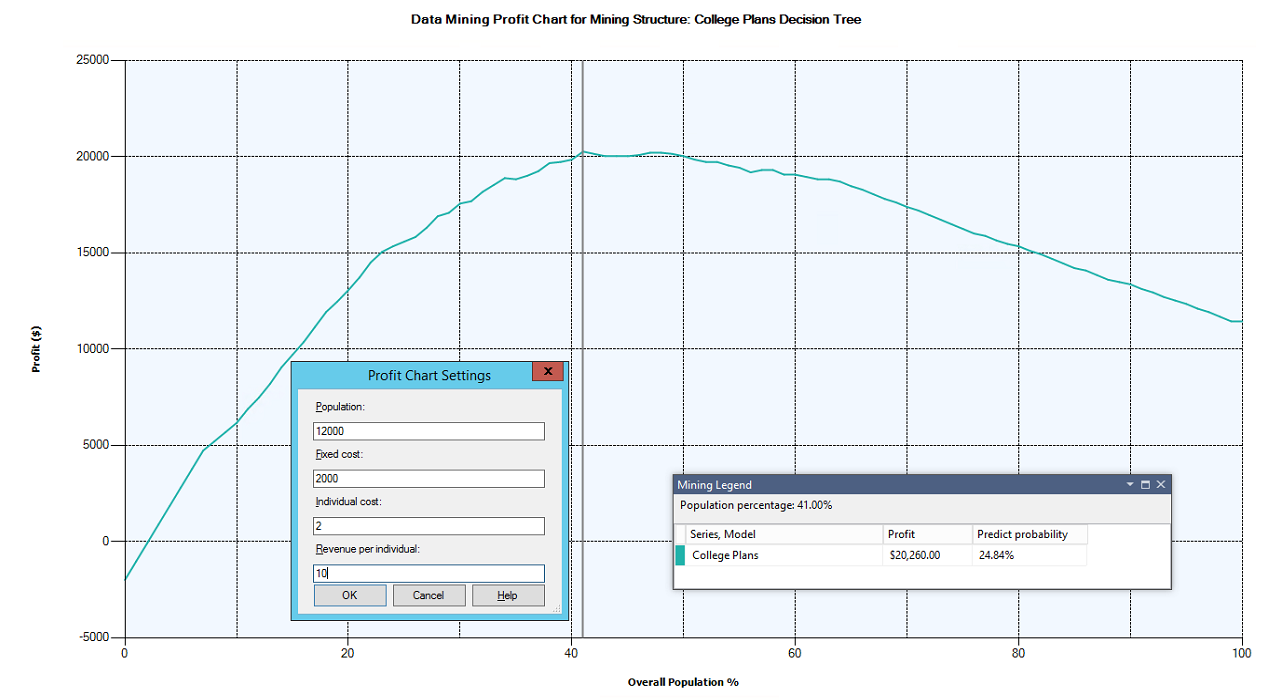
* + - 1. **Profit would be: 3,338 \* $10 = $33,380**

Support Diagrams

* + - 1. Decision Tree

Max Profit at 41% of population -> $20,260.00

Number of students -> 0.41 \* 12,000 = 4920



* + - 1. Neural Network

Max Profit at 52% of population -> $19,780.00

Number of students -> 0.52 \* 12,000 = 6240

