CIS 570 – Final Exam
Summer 2018

You should complete this exam on your own without assistance from anyone else. The deadline for submitting the exam is 11:59 PM on Wednesday, July 11.

Part A (36 Points)

On the SQL server instance, buscissql\cisbi, there is a database, ExamData, with a table, CollegePlans. The data in the table is for a sample of high school students, and includes their gender, IQ, parental income and encouragement, and whether they plan to attend college or not.

1. How many students are in the sample data set? What percentage of them have plans to attend college, and what percentage do not have plans to attend college? (1 point)

2. How many male students are in the sample? What percentage of the female students have plans to attend college? (1 point)

3. What is the mean (average) IQ of the students in the sample dataset? If a student's IQ is less than 90, is he/she more or less likely to have plans to attend college? Offer support for your answer. (2 points)

4. What is the minimum parental income in the sample dataset? 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)

5. Is Parental Encouragement likely to be a good predictor of a student's plans to attend college or not? Explain. (2 points)

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. Build, process and analyze the models within your analysis services database that was created for the midterm exam (**your Last Name + MidtermASD**). Reserve 25% of the data for testing and set the "HoldoutSeed" property to 75.

6. 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)

7. 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 b) not attend college. What are the probabilities for the two outcomes? (3 points)

8. 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)

9. 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)

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

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

Predicted	Does not plan to attend (Actual)	Plans to attend (Actual)
Does not plan to attend		
Plans to attend		

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

13. 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)

14. 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)

15. 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)

Submit the answers for the 15 questions in a Word or PDF file.

Part B (24 Points)

Use Power BI Desktop to create a minimum of three visualizations based on one or more data sets of your choice. This site has links to a few public datasets –

http://www.jenunderwood.com/2016/01/14/my-favorite-public-data-sources

Include a slicer in at least one of the visualizations. Add your name to the titles of each visualization. Format (e.g., color, font, background, border, positioning) the visualizations as you see fit. Custom visuals are available in the Microsoft Office store -

https://tinyurl.com/y7qwzjr2

Save your Power BI file in the PowerBI folder that you created earlier for the hands on exercises. Give the file a suitable name (e.g., SmithJPowerBIExam).

Use screen capture to prepare a document with the images of your visualizations.

Prepare another document with the following information:

- A description of each dataset including its source, number of rows, columns, and its relevance to you (i.e., why did you choose this dataset).
- The list of transformations (if any), that you performed on each of the datasets.
- For each visualization, provide a brief narrative that describes its purpose and the insight(s) it offers.

Submit the two documents in Word or PDF format.