

## CSCI 5080 Assignment 6

Use Microsoft **WORD** to write your answer and submit it to the Dropbox in D2L.

Please indicate **how much time** you spend on each problem.

### Part I (40 points)

Given the set of the following examples as a training set to build a decision tree:

Example	height	hair	eyes	Class
X1	short	dark	blue	+
X2	short	blond	blue	-
X3	tall	gray	brown	-
X4	tall	dark	blue	+
X5	short	dark	brown	-
X6	tall	blond	blue	+

1) (15 Points) Which attribute should be split on first? Why?

2) (20 Points) Show information gain computations that you used to induce a complete decision tree and draw the tree.

3) (5 Points) Give the class labels for the following new examples base on your decision tree.

Example	height	hair	eyes	Class
X7	short	gray	brown	
X8	tall	dark	brown	

## Part II (60 Points)

1. Review [SQL Server Analysis Services - Data Mining](http://technet.microsoft.com/en-us/library/bb677206(v=sql.110).aspx).  
[http://technet.microsoft.com/en-us/library/bb677206\(v=sql.110\).aspx](http://technet.microsoft.com/en-us/library/bb677206(v=sql.110).aspx)
2. Review SQL Server [Data Mining Algorithms](http://technet.microsoft.com/en-us/library/ms175595(v=sql.110).aspx).  
[http://technet.microsoft.com/en-us/library/ms175595\(v=sql.110\).aspx](http://technet.microsoft.com/en-us/library/ms175595(v=sql.110).aspx)

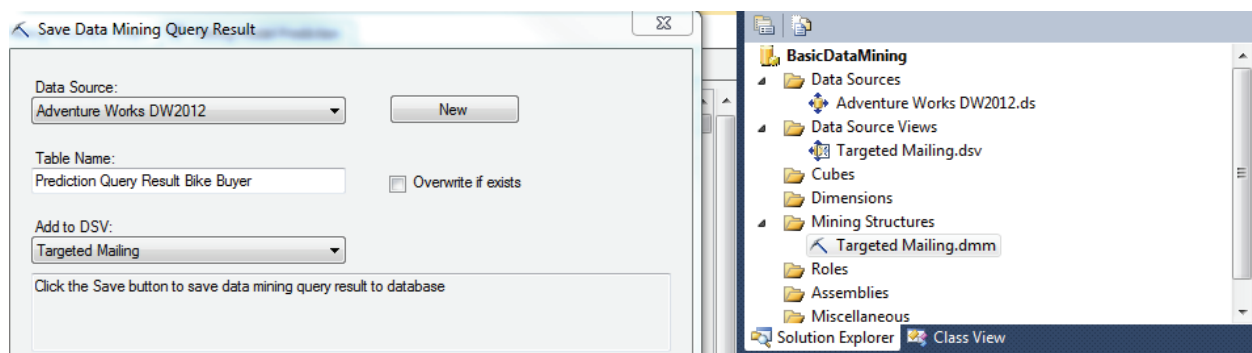
Study Microsoft [Decision Trees Algorithm](http://technet.microsoft.com/en-us/library/ms175312(v=sql.110).aspx)  
[http://technet.microsoft.com/en-us/library/ms175312\(v=sql.110\).aspx](http://technet.microsoft.com/en-us/library/ms175312(v=sql.110).aspx)

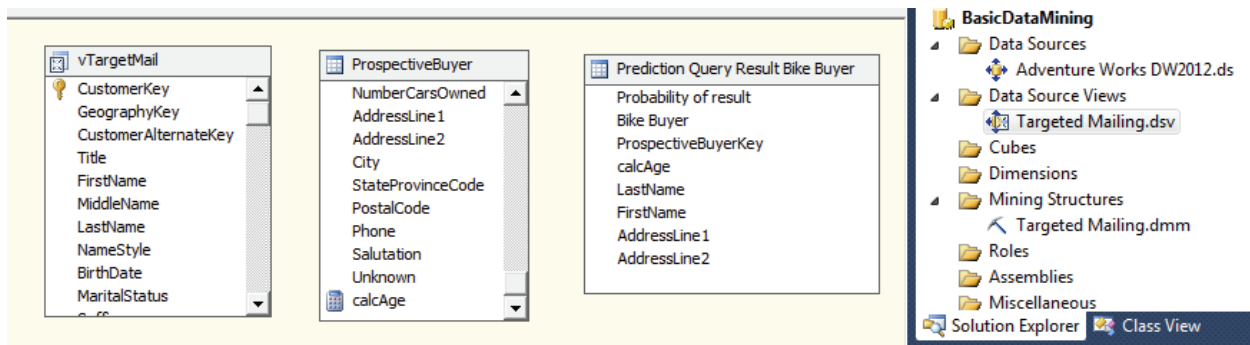
and Microsoft [Naive Bayes Algorithm](http://technet.microsoft.com/en-us/library/ms174806(v=sql.110).aspx)  
[http://technet.microsoft.com/en-us/library/ms174806\(v=sql.110\).aspx](http://technet.microsoft.com/en-us/library/ms174806(v=sql.110).aspx)

3. Complete [SQL Server 2012 Basic Data Mining Tutorial](http://technet.microsoft.com/en-us/library/ms167167(v=sql.110).aspx) Lesson 1 through Lesson 6.  
[http://technet.microsoft.com/en-us/library/ms167167\(v=sql.110\).aspx](http://technet.microsoft.com/en-us/library/ms167167(v=sql.110).aspx)

This tutorial is also available in the handout on SQL Server 2012 Tutorials: Analysis Services - Data Mining. You may skip the Clustering Model which will be discussed in Chapter 10 and skip the Drillthrough Queries in Lesson 6.

- (1) Turn in a completed project for each lesson, i.e. submit six completed projects for six lessons.
- (2) For lesson 4, turn in screenshots of the Decision Tree tab and the Dependency Network tab for the Decision Tree model, and the Attribute Characteristics tab and Attribute Discrimination tab for the Naïve Bayes model.
- (3) For lesson 5, turn in the screenshot of the Lift Chart that includes three Decision Tree models.
- (4) For lesson 6, save the prediction query result with the option Add to DSV “Targeted Mailing” with name “Prediction Query Result Bike Buyer”.





Note: You might need to **change server names**. The default server name is `localhost`. If your servers are installed on remote computers or as named instances, you must override the default to use a server name that is valid for your installation. Furthermore, if the servers are on remote computers, you might need to configure Windows Firewall to allow access to the servers.

For example, the server name in the screenshot below is `IT103949D\SQL2012` given the Instance ID `SQL2012` and my computer name `IT103949D`. The server name is specified in the Deployment tab of the Property Pages of the project.

