**CIS 8695 Big Data Analytics**

**Final Exam Study Guide**

**Exam Format: Closed book (one page (one sided) cheat sheet) Theoretical Questions (on paper) & Programming on your laptop; 2 hours and 15 minutes.**

**1. Linear Regression:** Chapter 6

* Understand dependent variables and independent variables;
* Understand estimated coefficients in the linear regression, and how to use them to make prediction;
* Understand how to interpret the coefficients;
* Understand what *R2*, *Adjusted R2* mean, and RMSE, and how they can be used to compare between different models;
* Understand the *three popular variable selection approaches:* forward, backward, and stepwise.

**2. Logistic Regression:** Chapter 10

* Understand and remember the model format of logistic regression, and what are *logit,* *odds, and predicted probability*;
* Understand how to interpret the meanings of coefficient estimates, how to **calculate predicted probability using coefficient estimates**, and how to classify the results based on the predicted probability (using the cutoff level).

**3. Classification Tree:** Chapter 9

* Understand how to interpret the results of a classification tree output, such as tree structure, the predictors used and the associated threshold levels, and classification rules;
* Understand how to use classification rules to classify (predict) new observations;
* Basic idea of random forest and variable importance in random forest;

**4. Naïve Bayes Classifier:** Chapter 8

* Understand the meaning of conditional probabilities from the Naïve Bayes model output;
* Understand how to use Naïve Bayes for classification.

**5. Performance Evaluation:** Chapter 5

* Understand confusion matrix, and concepts of **True Positive (TP), True Negative (TN), False Positive (FP), False Negative (FN)** used to evaluate the performance of model prediction. Understand their meanings.
* Understand ROC curve and the area under the curve and how we use them to evaluate models

**6. Dimension Reduction:** Chapter 4

* Understand what are *principal components*, and how they are related to the variances of original data values;
* Understand how to interpret the outputs of PCA, refer to page 105-109 of the textbook.

**7. Association Rules and Recommendation:** Chapter 14

* Understand the ***if-then*** format of association rules;
* - Understand the concepts of antecedent, consequent, support, confidence, benchmark confidence, and lift ratio;
* Understand how to interpret results of association rules analysis;
* Understand the main ideas of different collaborative filtering approaches.

**8. Neural Networks**

**-** Understand the general idea of neural network and how it works for prediction

**9. Time Series Forecasting**

- Understand the general idea of time series forecasting and its applications