**Machine Learning - CSCE 623 Spring 2019: In Class Work, Day 2 (28 Mar 2019)**

Part 1 (do this before our lecture):

1. In your own words, identify what you consider the most important aspect/activity for each step in the CRISP-DM process and explain why the step is important for the process as a whole:

Business Understanding:

Understand the purpose of the organization/role that the “business” plays and what it is trying to accomplish. Determine what it is you want to accomplish

Data Understanding:

Visualizing the data and figuring out what it initially looks like

Data Preparation:

Setting up the data to be fed into the model. Figuring out what data is going into the model

Modeling:

Figuring out how to test the model to ensure quality

Evaluation:

Assessing the results of the model. Did you get what you want?

Deployment:

Deployment plan, how are we going to put this out?

1. In your own words, identify what you consider the most important aspect/activity for each step in the KDD process (as described in “The KDD Process for Extracting Useful Knowledge from Volumes of Data”) and explain why it is important for the process as a whole:

Selection:

What data are we going to use? Important because if we use the wrong data, we get garbage results

Preprocessing:

How is our data formatted and what does it look like? Important for ease of input into model.

Transformation:

Reducing the number of variables of interest. Important because this affects how the model runs.

Data Mining:

Choosing the algorithms for data mining. Important because the choice of algorithm affects the results. If you need accuracy you would want to pick the more accurate algorithm for example.

Interpretation/Evaluation:

Correctly interpreting the output. Important because incorrectly doing this can derail the process.

Part 2 (to be completed during lecture when instructed):

1. For one of the following companies and describe how the company might use the modified KDD process (as described in class):

* NETFLIX
* Amazon
* Blizzard (computer gaming)
* USAA Insurance
* Delta Airlines

1. How do you envision using the modified KDD process (as described in class) in your AFIT research? What portion of your AFIT research could you work on for your class project?

