Florida Atlantic University
COLLEGE OF BUSINESS

ISM 6136 – Datamining/Predictive Analytics

Class Assignment 3
5 points

TASK: Performing classification using the k-NN classifier.

1. Download the Iris file provided and follow instructions below. The Iris dataset has data for four characteristics (sepal width, sepal length, petal width and petal length) of three species of Iris flowers. Predict which of the three species would the 'new data' flower fall into.

2. The target class/outcome is Species_name which has 3 classes. Predict the outcome for the new data provided in the spreadsheet.

3. Follow all the data mining steps:
   a. Understand the problem and purpose of data mining task
   b. Obtain the dataset for analysis
   c. Explore, clean and preprocess data
      - *Cleanup any column that is not a predictor*
      - *Perform 'Missing Data Handling' (For any missing data add 'Mean' instead of deleting for predictors and 'Delete' for outcome)*
   d. Determine the appropriate data mining task
   e. Partition data (supervised tasks)
   f. Choose the data mining techniques/algorithms
   g. Build the model by interpreting results of algorithms
   h. Compare the models and select the best one
   i. Deploy 'only' the best model (score on new data)

4. Try at least 5 models (different k values, partitioning, any other criteria) and choose the best one. (Select k values greater than 1).
5. **Provide explanation of steps c) through h) on a word document.** Regarding the model selection explain the 'full' model selection criteria (Slide# 16 of Lecture 5) you followed and how you compared each of the models. You can present it in tabular form too with results from each model.

6. Upload the Excel file and word document – showing all 4 models you tried and highlight the tabs of the best one.