

PESIT Department of Computer Science and Engineering

Course: Data Mining
Semester: 2016 Spring (January – May)
Instructor: BNR (Dr. B. Narsing Rao)

Assignment: 07
Topic: Classification – Classifier Comparison
Due by: Midnight on Tuesday, March 1, 2016
Method: See below for details; email to bnrao@pes.edu

Use the accompanying file **diabetes.arff** for this assignment.

Use the Weka GUI to classify the training set (i.e. predict the class attribute which indicates the presence or absence of diabetes, using 10 fold cross-validation) using the following methods:

- Decision Tree (J48)
- Naïve Bayes
- JRip
- Random Forest (in this case try using a different number of features and compare them)

Submit a PDF document (using the standard naming convention) with the following:

A table containing the comparison between these classification models (arranged in order of what you would consider to be the best to the worst) on the following parameters:

- Accuracy, Kappa Statistic, TP Rate, FP Rate, Precision, Recall, F Measure

Your observations and comments based on the results.