Course Project

Project steps:

1. Datasets:
   You can use the given datasets or your own datasets.
2. Experiments:
   a. explore different ways of discretizing continuous attributes. That is, convert numeric attributes into "nominal" ones by binning numeric values into intervals; explore different ways of removing missing values
   b. Use different ways of testing your results. For example, if you use supervised learning, explore the following alternatives:
      i. Testing your results over the training data.
      ii. Splitting your input file into two parts one for training and one for testing.
      iii. Using k-fold cross validation. Play with different values for k.
   c. Analyze the results obtained (i.e. interpret the meaning of the output).
3. Write a 2 to 4-page report of your analyses.
   a. Data: Describe the datasets that you used in terms of the attributes present in the data, the number of instances, missing values, and other relevant characteristics.
   b. Code Description: Describe to the extent possible any observations you made
   c. Experiments: For each experiment you ran describe:
      i. Instances: What data did you use for the experiments? That is, did you use the entire dataset or just a subset of it?
   d. Any pre-processing done to the data. That is, did you remove any attributes? Did you discretize any continuous attribute? If so, what strategy did you use to bin the values? Did you replace missing values? If so, what strategy did you use to select a replacement of the missing values?
   e. Analysis of the experiments' results
   f. Summary of Results
   g. Discuss the strengths and the weaknesses of your project.