Data Mining Project 2 Grading Policy

If you have any questions, please e-mail to nckudm@gmail.com.

1. Topic: Classification Analysis

2. Data (35 pts)
   You should design your own data; using data online will result in 0 points in this section.

   - Problem Definition / Data Design (10 pts)
   - 10 Features (10 pts), you can add redundant features and observe the performance.
   - 5 Rules (15 pts)

3. Classification Models (25 pts)

   - Decision Tree (10 pts)
   - Any models of your preference (15 pts), eg. SVM, KNN, Bayes, …

4. Report (40 pts)

   - Decision Trees (10 pts)
     Include your decision tree figure(s).
   - Comparisons (15 pts)
     Compare your absolutely right rules with the rules generated by the classification model(s).
   - Discussion (15 pts)
     Slightly alter the absolutely-right rules and generate another set of data; run the classifcation models on this set of data and include your observations in this section.

5. Submission Format
   Please submit a zipped directory (<= 200 MB) with the below structure:

   ```
   hw2
   ├── inputs (directory for input files)
   │   ├── data.csv // your generated data; you can use any names you like with fil
   │   ├── data2.csv // include other data if you have more
   ├── main.py      // the code you use to run the classification models
   ...              // include other scripts if you have more
   └── report.pdf   // your report file, be sure it is a .pdf
   ```

   TA will not run your code for this project, but please make sure
   that you hand in the code that produces the input data and executes the classification
   and try to make it readable with comments, lest we would need to refer to it under any
   circumstances.

6. Submission Deadline: **11/22 (Tue.) 9:00 a.m.**

   - Note that the deadline is 9:00 in the morning!
   - Late submissions within 2 days (before 11/24 (Thu.) 9:00 a.m.) will receive a 20% discount on the overall score.
   - Submissions delayed more than 2 days will not be graded.

7. [Past Works For Your Reference](#)