

Project 2: Classification

UE 141 on Data Mining Spring 2013

Each team needs to present project in class and submit a written report before class on April 10. The maximum number of members in a team is 2.

The goal of this project is to evaluate your understanding of classification in real-world applications, and get you familiar with the classification algorithms in Weka.

You are asked to complete the following two tasks.

- Please discuss classification in the context of a real-world application. What are the data sets? What is the expected output? How can classification help decision making in this application?
- Conduct classification on a data set. Run the J48 algorithm (decision tree) on the data. How does the tree look like? What is its classification accuracy? Run IB1 (nearest neighbor) on the data. Does it have a classifier? What is its classification accuracy?

Your report should include: 1) A short description of the application and answers to the three questions in part 1, and 2) A short description of the data set you use for Weka experiments. Answer the questions in part 2.

Your presentation should be a short overview of your report. You can ignore the details but show the most interesting points.

Note that plagiarism/copying is not allowed and may result in an F in the grades of all the team members. Academic integrity policy can be found at <http://www.cse.buffalo.edu/shared/policies/academic.php>