

INFORMATION DISCOVERY FROM CANADA THISTLE CONTROL RESEARCH DATA BY USING CLASSIFICATION MINING. Jingkai Zhou, Janet Davidson-Harrington, and Calvin G. Messersmith, PostDoc, Research Specialist, and Professor, Department of Plant Sciences, North Dakota State University, Fargo, ND 58105

Data-mining is a technique that extracts hidden predictive information that experts may miss because it lies outside their expectation. Classification mining, which is a data-mining task, aims to identify characteristics that indicate the group to which each case belongs. The classifier (model) from classification mining can be used both to understand existing data and to predict future trends and behaviors, which allows users to make proactive, knowledge-driven decisions. A classification mining function has been developed based on Canada thistle control research data, but this function with limited modification can mine other databases. One goal of classification mining functions is to calculate the entropy (data stability) for each predictive attribute, such as temperature before or after treatment. The predictive factor(s) with the lowest entropy is the factor(s) that has the most effect on the predicted factor, such as control. For example, using 2,4-D for Canada thistle control the entropy of minimum and maximum temperature at 1 d after 2,4-D treatment is 0.901 and 0.857, respectively, indicating that maximum temperature after treatment has more effect than minimum temperature on Canada thistle control by 2,4-D. Another goal is create a classifier. First, a classifier, i.e., a decision tree, is built that describes a predetermined set of data classes or concepts. When the accuracy of the classifier is considered acceptable, the classifier can be used to classify future data.