

Data Analytics 874  
Post-Block Assignment 2: Rule Extraction and Association  
Rule Mining  
Department of Industrial Engineering

Deadline: 15 January 2021, 23:59

## Instructions

For the purposes of this assignment, note the following instructions:

- Assignments are to be completed by students individually.
- Submit your own work.
- You will submit a pdf document addressing all of the aspects listed in the specifications below. Name your pdf document ????????PBA2\_874.pdf, where the question marks are replaced with your student number.
- Make sure that your pdf document has a title page containing the assignment number as the title, and your initials, surname and student number.
- Submit your pdf document no later than the deadline.

## Rule Extraction

This assignment continues with the analysis of the abalone dataset. For the purposes of this assignment, use the cleaned abalone dataset as provided for post-block assignment 1. For this part of the assignment, you will extract rules from the abalone dataset, using the number of rings as the target feature.

The assignment will be completed in the following steps:

1. The first thing that you have to decide is if you will treat this problem as a regression problem, or as a multi-class classification problem. The nature of this dataset is such that it can be considered as any of these two problem types. In your report clearly state your selection of problem type. (5)
2. Based on whether you treat this as a regression or classification problem, you will first induce a decision tree for the abalone dataset. Choose an appropriate decision tree. In your report give a short description of your choice of decision tree with a short motivation. Then proceed to quantify the prediction performance of the decision tree induced on the abalone dataset. If you have done any data transformations or data processing prior to induction of the decision tree, please describe these with clear motivations. (20)

3. For the next step, extract rules in the form **if antecedent then consequent** from the induced decision tree. Quantify the performance of each rule in terms of its coverage and its accuracy (both with respect to accuracy on the training set and generalization performance). Evaluate the performance of the entire rule set, and compare that with the performance of the decision tree. In addition, discuss the complexity of the rules extracted. (15)
4. Now, apply a pruning process on the rule set, and evaluate the resulting rules with reference to coverage, accuracy and complexity. In your report, also describe the rule pruning process that you have applied. (15)
5. Decide on a covering rule induction algorithm and extract a set of rules from the abalone dataset. Describe the rule induction algorithm that you have selected, and evaluate the extracted rule set. (30)
6. Use the same rule pruning approach as above, and evaluate the performance of the extracted rules. Compare these with the unpruned rule set, and draw conclusions on whether rule pruning has played any role in improving the quality of the extracted rule set. (15)
7. Lastly, proceed to compare the rule sets extracted from the decision tree and from the rule induction algorithm. (15)

## Rule Association

For this part of the assignment, you will apply a rule association algorithm to the abalone dataset to explore any additional patterns from the dataset. Select an appropriate rule association algorithm, and provide a short description. State all of the control parameters of the algorithm and provide the values used for these control parameters with motivations. Then proceed to report the association rules that you have discovered. Provide an analysis of the performance of these association rules. Proceed to select only those association rules that include the target feature, and compare the performance of these rules with that extracted using the rule induction algorithm above. (30)