The University of Texas at Austin
CE 395 R 5-Data Mining – Spring 2018

Homework 2 - Individual Assignment

DUE DATE: 03/02/2018

Please submit the assignment by the beginning of the lecture.

PROBLEM 1: Visit the UCI ML Repository. Download the "Glass Identification Database" data file (glass.data). Open the data base from the WEKA Software Explorer GUI and perform the following data preprocessing tasks:

- Remove the attribute "ID number".
- Normalize all numerical attributes in the data set to a 0 to 1 range.
- Discretize the numerical attribute "Mg: Magnesium" into 8 bins.

Save the changes to a file containing the modified data set in ARFF format.

PROBLEM 2: Build a Decision Tree classification model for the following data set. For this problem, the students should NOT use a computer-based decision tree algorithm. All calculations should be done by hand or using a spreadsheet program.

record	alternative	bar	Friday	hungry	customers	price	rain	reservation	type	wait	go/no go
x1	yes	no	no	yes	some	$$$	no	yes	french	0-10	yes.
x2	yes	no	no	yes	full	$	no	no	thai	30-60	no.
x3	no	yes	no	no	some	$	no	no	burger	0-10	yes.
x4	yes	no	yes	yes	full	$	no	no	thai	10-30	yes.
x5	yes	no	yes	no	full	$$$	no	yes	french	>60	no.
x6	no	yes	no	yes	?	$$	yes	yes	italian	0-10	yes.
x7	no	yes	no	no	none	$	yes	no	burger	0-10	no.
x8	no	no	no	yes	some	$$	yes	yes	thai	0-10	yes.
x9	no	yes	yes	no	full	$	yes	no	burger	>60	no.
x10	yes	yes	yes	yes	full	$$$	no	yes	italian	10-30	no.
x11	no	no	no	no	none	$	no	no	thai	0-10	no.
x12	yes	yes	yes	yes	full	$	no	no	burger	30-60	yes.

Legend:
alternative: Is there an alternate restauraunt?
bar: Does the restaurant have a bar?
Friday: Is it Friday?
hungry: Are you hungry?
customers: Are there customers at the restaurant?
price: What is the price range?
rain: Is it raining?
reservation: Do you have a reservation?
type: What is the restaurant type?
wait: what is the waiting period?
?: missing data

SUBMISSION: The following files should be submitted:

- File 1 (Problem 1): the modified "Glass Identification Database" in ARFF format.
- File 2 (Problem 2): one file (in Word, Excel, or PDF format) describing all steps and the final decision tree model.