NATIONAL UNIVERSITY OF SINGAPORE

## ST5227 Applied Data Mining

(Semester 2 : AY 2019/2020)

Final Assessment

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

# INSTRUCTIONS TO CANDIDATES

1. Please select two data mining methods that you have learned in this module and conduct the following analysis;

2. Give a detailed introduction on the two methods, clearly define all the mathematical notations used in your introduction including variables, parameters, distributions, equations and models etc.

3. Use one data set to illustrate the two methods and compare their performances, you **cannot** use any of the data sets used in our lectures or tutorials; describe the basic data information and interpret your results; include the data source (website or package), detailed R code and computer output and refer to them appropriately in your project.

4. The two selected methods should be applicable for the same kind of data and so are comparable. For example, you can select KNN and CART for classification, or spline regression and local linear regression; but you cannot select multi-level logistic regression and binary response support vector machine.

5. All methods **except** linear regression estimated by least squares and binary logistic regression model can be selected.

6. You are NOT allowed to **communicate** with each other. Any submissions with high degree of similarity in contents will be subjected to plagiarism penalty.

7. Maximum page limit is 8, including all references, tables and graphs (if you have any). Type your matric number on the first page of your submitted report. Use Arial 12 fonts with the same paragraph style as in this instruction.

8. Submit your report to Luminus folder “Final submission” by 19:00 on 1st May. Only DOC and PDF files are acceptable for this submission. Do not submit Zipped files or multiple files.