

 UNIVERSITI TEKNIKAL MALAYSIA MELAKA	<b>Case Study: Exploratory Analysis</b> FACULTY OF INFORMATION & COMMUNICATION TECHNOLOGY UNIVERSITI TEKNIKAL MALAYSIA MELAKA	
<b>MACHINE LEARNING</b>		
<b>BITI 2223</b>	<b>SEMESTER 2</b>	<b>SESI 2023/2024</b>

This assignment is to be completed in groups of two members each.

**Instructions:**

- 1) You are required to examine **ONE** of the following data sets (based on group) from the UCI Machine Learning Repository (<https://archive.ics.uci.edu/>) using the RapidMiner Studio:

1. Travel Review Ratings Data Set
2. EMG data for gestures Data Set
3. Student Academics Performance Data Set
4. EEG Steady-State Visual Evoked Potential Signals Data Set
5. Cargo 2000 Freight Tracking and Tracing Data Set
6. Breast Cancer Wisconsin (Diagnostic)
7. Drug Consumption (Quantified)
8. Myocardial infarction complications
9. Cervical Cancer (Risk Factors)
10. Bar Crawl: Detecting Heavy Drinking
11. In-Vehicle Coupon Recommendation

- 2) You are required to prepare a report of **maximum 5 pages** including all figures for this assignment by answering the following questions. An additional 5 pages is allowed for attachment.

**1. Data Description**

- a. Name of dataset
- b. Brief description
- c. Source of data
- d. List of attributes with their types (categorical, nominal and continuous).

**2. Give 3 business questions that can be asked on this data set, for each question:**

- a. What do you expect to discover from mining on this data?
- b. What types of relationships are being mined (classification rules, clusters, tree structure, association rules, etc.)?

**3. Preliminary examination of data**

- a. Are there any “obvious” patterns in the data which might be helpful? (By observing the data in spread sheet and perform initial statistical check such as min/mode/histogram/median/standard deviation, correlation analysis etc.)
- b. State at descriptive analysis methods according to the *Roadmap for Data Exploration*. Justify your suggestion. Give the purpose and the hypothesis for each of the suggested methods.

**DUE DATE OF SUBMISSION: 28/5/2024 (Tuesday)**

**SUBMIT THROUGH: ULEARN**

<b>Evaluation:</b>	<b>Marks</b>
Satisfying the requirement of the questions	5
Give reasonable justification on the answer	5