



UNIVERSITI TUN HUSSEIN ONN MALAYSIA

**FINAL EXAMINATION
SEMESTER I
SESSION 2019/2020**

COURSE NAME : DATA SCIENCE
COURSE CODE : BWB 20403
PROGRAMME CODE : BWQ
EXAMINATION DATE : DECEMBER 2019/ JANUARY 2020
DURATION : 3 HOURS
INSTRUCTION : ANSWER ALL QUESTIONS

TERBUKA

THIS QUESTION PAPER CONSISTS OF **THREE (3)** PAGES

TERBUKA

- Q1** (a) Differentiate between single imputation and multiple imputation. (6 marks)
- (b) Illustrate **THREE (3)** types of missing data mechanism. (9 marks)
- (c) Discuss any **THREE (3)** methods for handling missing data. (6 marks)
- Q2** (a) Define the meaning of Data Science. (2 marks)
- (b) Discuss **FIVE (5)** issues in data management. (10 marks)
- (c) Outline **SIX (6)** steps in data life cycle. (12 marks)
- Q3** (a) Define the meaning of Big Data. (3 marks)
- (b) List down and explain **THREE (3)** major issues in handling Big Data. (6 marks)
- (c) Discuss **FOUR (4)** challenges of using Hadoop in Big Data analysis. (8 marks)
- Q4** (a) Define the meaning of outliers. (2 marks)
- (b) Explain **THREE (3)** tests for outlier detection in regression analysis. (6 marks)
- (c) Identify the outliers for data in **Table Q4(c)** by constructing a box and whisker plot and plot the data set.

Table Q4(c)

30	21.4	39.4	33.4	21.1	29.3	32.8	31.6	36
27.9	27.3	29.4	29.1	38.6	33.8	29.1	37.1	

(12 marks)

TERBUKA

- Q5** (a) Explain Monte Carlo simulation and how it works. (6 marks)
- (b) Explain the meaning of statistical pattern recognition. (4 marks)
- (c) Discuss any **TWO (2)** basic models in pattern recognition. (8 marks)

– **END OF QUESTIONS** –

TERBUKA