UNIVERSITI TEKNOLOGI PETRONAS

FINAL EXAMINATION
JANUARY 2025 SEMESTER

COURSE : TEB2043 - DATA SCIENCE

DATE : 11 APRIL 2025 (FRIDAY)

TIME : 9:00 AM - 12:00 NOON (3 HOURS)

INSTRUCTIONS TO CANDIDATES

1. Answer **ALL** questions in the Answer Booklet.
2. Begin **EACH** answer on a new page in the Answer Booklet.
3. Indicate clearly answers that are cancelled, if any.
4. Where applicable, show clearly steps taken in arriving at the solutions and indicate **ALL** assumptions, if any.
5. **DO NOT** open this Question Booklet until instructed.

Note :

i. There are **SIX (6)** pages in this Question Booklet including the cover page .

ii. **DOUBLE-SIDED** Question Booklet.

1. A healthcare organization is leveraging data science to improve patient outcomes, streamline hospital operations, and enhance disease prediction.

   a. Explain **THREE (3)** types of Data Science that can be used in healthcare and provide relevant examples.

   [6 marks]

   b. Identify **FOUR (4)** key steps in understanding a healthcare-related business problem before applying data science techniques.

   [4 marks]

   c. Discuss **FIVE (5)** key characteristics of Big Data and use real-world examples related to healthcare.

   [10 marks]

2. a. Discuss **FOUR (4)** importance of Data Quality Management (DQM) in a data-driven organization.

[8 marks]

b. A national retail chain has recently adopted a data-driven approach to manage inventory, track customer preferences, and optimize sales strategies. However, the company is struggling with data quality issues, leading to inaccurate stock levels, misleading customer insights, and ineffective marketing campaigns. As a result, stores frequently experience stockouts of popular products, overstocking slow-moving items, and incorrect customer segmentation, which affects personalized promotions. Discuss **FOUR (4)** causes of data quality issues in the retail chain and analyze their impact on decision-making. Support your discussion with examples.

[12 marks]

3. a. Beautiful Sdn. Bhd. wants to launch a new skincare product and needs customer insights. Apply your knowledge of primary and secondary data by suggesting the **THREE (3)** suitable type for this scenario and explaining how it would be collected and used.

[6 marks]

b. Beautiful Sdn. Bhd. is facing issues with inaccurate records in its database. Discuss the significance of data cleansing in resolving these issues and explain **FIVE (5)** key steps to enhance the data quality. Provide relevant examples to support your explanation.

[14 marks]

4. You are a data analyst at a research institute studying student health and wellness. You are given the following dataset representing the weight of 15 students, as shown in **TABLE Q4**.

TABLE Q4: Test Score

Student	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
Weight (kg)	55	60	65	50	70	55	62	68	53	61	58	66	72	57	64

a. Calculate the Measures of Central Tendency (mean, median, mode) for the students' weight.

[6 marks]

b. Calculate the range, variance, and standard deviation of the students' weight.

[6 marks]

c. Based on your calculations in (a) and (b), are the students' weight closely clustered around the mean or widely spread out? Justify your answer.

[4 marks]

d. Create a histogram of the students' weight and analyze its shape.

[4 marks]

5. A technology company is developing an AI-driven recommendation system for an e-commerce platform. The goal is to improve user experience by suggesting relevant products based on customer behavior and preferences.

a. Explain how supervised learning approach can be used to predict customer purchase behavior based on historical transaction data and propose an appropriate machine learning model for this task with justification.

[6 marks]

b. Explain how unsupervised learning can be applied to group customers with similar shopping habits for personalized recommendations and propose a suitable clustering algorithm with justification.

[6 marks]

c. Compare **TWO (2)** advantages and limitations of each learning method in this scenario.

[4 marks]

d. Recommend the most suitable learning approach for optimizing the recommendation system, providing justification.

[4 marks]

- END OF PAPER -