UNIVERSITI MALAYSIA PAHANG
AL-SULTAN ABDULLAH

FINAL EXAMINATION

COURSE	**:**	**ARTIFICIAL INTELLIGENCE TECHNIQUES**
COURSE CODE	**:**	**BCS2313**
LECTURER	**:**	**DR. NUR SHAZWANI BINTI KAMARUDIN**
DATE	**:**	**12 FEBRUARY 2025**
DURATION	**:**	**3 HOURS**
SESSION/SEMESTER	**:**	**SESSION 2024/2025 SEMESTER I**

INSTRUCTIONS TO CANDIDATE:

1. This question paper consists of **THREE (3)** questions. Answer **ALL** questions.
2. Write your answers in the answer booklet provided.
3. All answers to a new question should start on a new page.
4. All the calculations and assumptions must be clearly stated.
5. The question should be answered in English

EXAMINATION REQUIREMENTS:

1. Answer booklet
2. Calculator

APPENDIX:

1. Appendix 1 – Formula Sheet

DO NOT TURN THIS PAGE UNTIL YOU ARE TOLD TO DO SO

This examination paper consists of **SEVEN (7)** printed pages including the front page.

QUESTION 1 [43 MARKS]

As a software developer at Pekan Medical Centre, you are tasked with designing, implementing, and evaluating a comprehensive Case-Based Reasoning (CBR) system to support medical professionals in the diagnosis and treatment of heart disease. The goal of the system is to assist doctors by retrieving similar past cases from a case base to provide insights and recommendations for diagnosing new patients. These cases will include detailed information about patient symptoms, medical history, diagnostic test results, treatments administered, and patient outcomes. By comparing the new patient's data with similar historical cases, the system will help doctors make informed decisions about diagnosis, treatment options, and potential outcomes.

To develop this system, you will need to demonstrate a thorough understanding of the four primary components of CBR: case retrieval, where relevant past cases are identified based on their similarity to the new case; case reuse, where the system suggests solutions based on the retrieved cases; case revision, where the solution may be adapted to better fit the current situation; and case retention, where new cases are stored in the system to enhance its knowledge base for future use.

Table 1: **Cased-Based Reasoning Table**

Case ID	Age	Cholesterol Level (mg/dL)	Blood Pressure (mm Hg)	Heart Rate (bpm)	Family History (0 = No, 1 = Yes)	Outcome (0 = No Disease, 1 = Disease)
1	55	240	140	80	1	1
2	60	210	135	75	1	1
3	45	180	120	70	0	0
4	50	200	130	85	1	1
5	65	230	145	78	1	1

Table 2: New Case

Case ID	Age	Cholesterol Level (mg/dL)	Blood Pressure (mm Hg)	Heart Rate (bpm)	Family History (0 = No, 1 = Yes)	Outcome (0 = No Disease, 1 = Disease)
6	52	220	138	80	1	TBD

Explanation of Attributes:

- **Age**: The patient's age.
- **Cholesterol Level**: Measured in mg/dL; high cholesterol is a risk factor for heart disease.
- **Blood Pressure**: Measured in mm Hg; high blood pressure is a significant factor in diagnosing heart disease.
- **Heart Rate**: Measured in beats per minute (bpm); it reflects the patient's heart condition.
- **Family History**: A binary indicator showing whether the patient has a family history of heart disease (1 = Yes, 0 = No).
- **Outcome**: Indicates if the patient was diagnosed with heart disease (1 = Disease, 0 = No Disease). For the new case, this is not yet known (TBD).

(a) Describe Knowledge Discovery in Databases (KDD). Discuss the importance of KDD in data science.

[4 Marks]

[CO2, PO1, C2]

(b) What is data preprocessing in KDD and discuss its necessity.

[4 Marks]

[CO2, PO1, C2]

(c) Give an example of a real-world application of KDD, explain how the KDD process is applied.

[5 Marks]

[CO2, PO1, C3]

(d) Based on the case study, explain how Case-Based Reasoning (CBR) can improve over time.

[3 Marks]

[CO2, PO1, C4]

(e) Predict the outcome for **Case ID 6** and rounded your answer to **TWO (2)** decimal places.

[27 Marks]

[CO2, PO1, C4]

QUESTION 2 [40 MARKS]

Unity College is implementing an AI-driven system to predict students' final exam performance in a specific course. The system considers multiple factors, including midterm exam scores, attendance rates, and weekly study hours, as inputs to model the relationship with the final exam score. To achieve this, the college employs a simple neural network structure.

(a) The network uses the Mean Squared Error (MSE) loss function to measure prediction errors. Assuming the inputs and outputs are normalized between 0 and 1, visualize and draw the architecture of this neural network.

[6 Marks]

[CO2, PO1, C4]

(b) Using the following initial weights and biases in **Table 3**, perform the forward pass to calculate the output of the network for one data point with the input values:

Table 3

Inputs	$X_1 = 0.8$	$X_2 = 0.6$	$X_3 = 0.9$
Weights for input to hidden layer connections	$W_{1,1} = 0.2$ $W_{1,2} = -0.4$	$W_{2,1} = 0.3$ $W_{2,2} = 0.6$	$W_{3,1} = -0.5$ $W_{3,2} = 0.1$
Biases for hidden layer neurons	$b_1 = 0.1$	$b_2 = -0.2$	
Weights for hidden layer to output neuron connections	$W_{1,out} = 0.7$	$W_{2,out} = -0.3$	
Bias for output neuron	$b_{out} = 0.05$		

[9 Marks]

[CO2, PO1, C3]

(c) Given the true final exam score (normalized to 1) is 0.85, calculate the Mean Squared Error (MSE) loss between the predicted output from the network and the actual value. Show your calculations.

[5 Marks]

[CO2, PO1, C4]

(d) Assuming a learning rate of 0.01, calculate the partial derivatives of the MSE loss with respect to each weight between the hidden layer and the output layer only.

[8 Marks]

[CO2, PO1, C4]

(e) Discuss the importance of normalization of the inputs between 0 and 1 before feeding them into the neural network? Explain **TWO (2)** benefits of normalization to the training process.

[6 Marks]

[CO2, PO1, C3]

(f) Describe how the Mean Squared Error (MSE) loss function is used in this neural network model and explain **TWO (2)** reasons why MSE is an appropriate choice for this prediction task.

[6 Marks]

[CO2, PO1, C3]

QUESTION 3 [17 MARKS]

ParkFind, a theme park, wants to improve visitors' experience by helping them find the shortest walking path between popular attractions. The park is represented as a grid, where each node represents a location or intersection, and edges represent paths between them. Due to park layout and one-way pathways, some paths have different distances, and certain areas may be restricted or closed.

You have been tasked with designing a basic search algorithm to find the shortest route between two attractions, "Entrance" and "Roller Coaster." The park layout is shown in **Figure 1** below.

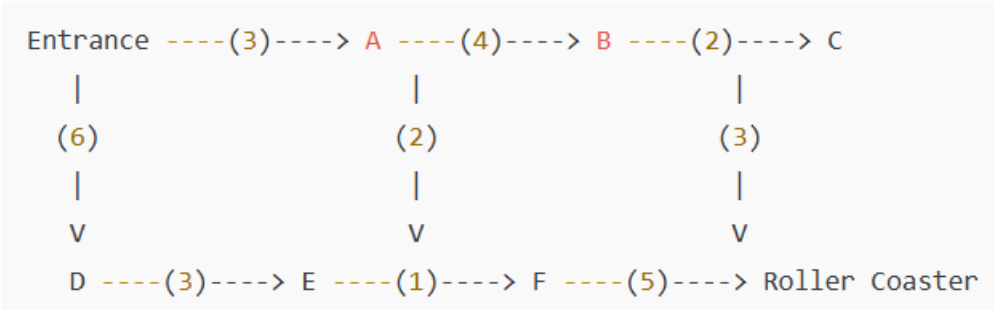

Figure 1

(a) Using a A* search algorithm, demonstrate how you would find the shortest path from "Entrance" to "Roller Coaster." Show each step of the process and calculate the total distance.

[10 Marks]

CO2, PO1, C4]

(b) If the path between D and E becomes temporarily blocked, explain how this change would impact the algorithm. Describe how you would modify the graph and recalculate the path.

[7 Marks]

[CO2, PO1, C4]

END OF QUESTION PAPER

APPENDIX 1 - Formula Sheet

Local Similarity	$LS = 1 - \frac{\lvert a - b \rvert}{max - min}$
Global Similarity	$sim(A,B) = \frac{1}{\sum w_i} \cdot \sum_{i=1}^{p} w_i \cdot sim_i(a,b)$
Fuzzy Sets (Intersection)	$\mu_{A \cap B}(x) = min[\mu_A(x), \mu_B(x)] = \mu_A(x) \cap \mu_B(x)$
Fuzzy Sets (Union)	$\mu_{A \cap B}(x) = max[\mu_A(x), \mu_B(x)] = \mu_A(x) \cup \mu_B(x)$
Min, Max Normalization	$v' = \frac{v - \min A}{\max A - \min A} (new\_\max A - new\_\min A) + new\_\min A$
Error	$error = (weight_k * error_j) * transfer_{derivative}(output)$