

33

**ISLAMIC UNIVERSITY OF TECHNOLOGY (IUT)**  
**ORGANISATION OF ISLAMIC COOPERATION (OIC)**  
**Department of Computer Science and Engineering (CSE)**

SEMESTER FINAL EXAMINATION  
 DURATION: 3 HOURS

WINTER SEMESTER, 2022-2023  
 FULL MARKS: 150

**CSE 4561: Computer Science and Technology III**

**Programmable calculators are not allowed. Do not write anything on the question paper.**  
 Answer all 6 (six) questions. Figures in the right margin indicate full marks of questions whereas corresponding CO and PO are written within parentheses.

1. You are a junior data analyst working in the marketing analyst team at Cyclistic, a bike-share company in Chicago. The director of marketing believes the company's future success depends on maximizing the number of annual memberships. Therefore, your team wants to understand how casual riders and annual members use Cyclistic bikes differently. After data collection, your team realized that the dataset was too large to be stored in traditional settings. The company also lacks the processing capabilities for such a large dataset. The team plans to gather insights and design a new marketing strategy to convert casual riders into annual members. But first, Cyclistic executives must approve your recommendations, so they must be backed up with compelling data insights and professional data visualizations.
- Based on the given scenario, answer the following questions:
- |                                                                                                                                                                                                                                                                                                                                                                                                                                                          |                         |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------|
| a) Write two relevant questions that will help you define the problem statement.                                                                                                                                                                                                                                                                                                                                                                         | 5 × 6                   |
| b) What steps should you take to plan and collect the data? Mention the tools.                                                                                                                                                                                                                                                                                                                                                                           | (CO2,<br>CO3)           |
| c) Based on the 3 V's, can you classify the dataset as big data?                                                                                                                                                                                                                                                                                                                                                                                         | (PO2)                   |
| d) Your dataset is tabular, contains multiple columns with null values, and has duplicate rows. How can you clean the dataset?                                                                                                                                                                                                                                                                                                                           |                         |
| e) What are the external services that you can rely on to analyze the dataset?                                                                                                                                                                                                                                                                                                                                                                           |                         |
| f) Assuming necessary data insights, mention two marketing strategies that can be presented to the executives.                                                                                                                                                                                                                                                                                                                                           |                         |
|                                                                                                                                                                                                                                                                                                                                                                                                                                                          |                         |
| 2. a) Define statistical modeling. Provide an example of a statistically insignificant event.                                                                                                                                                                                                                                                                                                                                                            | 3 + 4<br>(CO1)<br>(PO1) |
| b) You joined as a marketing manager for a large e-commerce company. Your role involves working with a large dataset of customer data containing their purchase history, demographics, and interests. Your goal is to use this data to segment your customers into different groups, enabling the company to target them with relevant marketing campaigns. With proper reasoning, propose a machine learning algorithm suitable for the given scenario. | 6<br>(CO3)<br>(PO2)     |
| c) You are an active reader who keeps track of the books in a mobile app called Goodreads. Recently, Goodreads launched an audiobook service to capitalize on its good reputation as a book cataloging app. Which AI system should be integrated with Goodreads to highlight audiobooks that the users might like?                                                                                                                                       | 4<br>(CO3)<br>(PO2)     |
| d) Provide example scenarios for the following topics:                                                                                                                                                                                                                                                                                                                                                                                                   | 4 × 3                   |
| i. Evolution of organizational structure in business                                                                                                                                                                                                                                                                                                                                                                                                     | (CO1)                   |
| ii. Cloud-based solutions providing Software-as-a-Service (SaaS)                                                                                                                                                                                                                                                                                                                                                                                         | (PO1)                   |
| iii. Data-driven decision outperforming gut instincts                                                                                                                                                                                                                                                                                                                                                                                                    |                         |

3. Your teacher is trying to predict the marks of his students based on their class attendance using machine learning. However, the teacher is new to machine learning and is intimidated by the complex algorithms. He would rather prefer a simple approach in this regard. Through a survey, he managed to create a small dataset. For the dataset mentioned in Table 1, answer the following questions.

**Table 1:** Student Dataset for Question 3

ID	Name	Attendance	Marks
123485	Farhan	80%	65
123456	Ishmam	85%	56

- a) Identify the Personal Identification Information (PII) in the dataset. 4  
(CO2)  
(PO2)
- b) How should he transform the dataset to ensure the privacy of the students? 3  
(CO1)  
(PO1)
- c) Which machine learning algorithm can help him in this regard? Justify your answer. 5  
(CO3)  
(PO2)
- d) Based on the given dataset, answer **True/False** for the following statements. If **False**, provide the correct answer. 1 × 6  
(CO1)  
(PO1)
- Attendance is a form of qualitative data.
  - ID is a form of ordinal data.
  - None of the columns in the dataset is an example of unstructured data.
  - The dataset is an example of external data.
  - The dataset cannot be classified as big data primarily due to the lack of variety in the dataset.
  - The dataset should be stored in transactional databases.
4. The multinational corporation, RiseTech, specializing in financial services gathers extensive customer data of millions of users to analyze spending patterns and provide personalized financial advice. The data used for analysis is stored in database A which exceeds a few hundred Gigabytes and is rarely updated. The customer data gathered by RiseTech is sourced from a separate database B that keeps track of purchases of millions of users on a daily basis. The associated dataset exceeds a few hundred Terabytes with a few columns containing Customer ID, Product ID, quantity, and price.
- a) Identify which criteria of big data are fulfilled by databases A and B respectively. Can the datasets be classified as big data? 5 × 2  
(CO2)  
(PO2)
- b) Which type of database is appropriate for databases A and B respectively? Justify your answer. 4 × 2  
(CO3)  
(PO2)
- c) Customer spending pattern is sensitive data and requires a secure tunnel between your computer and the database server. Which technology can be used to ensure such security? 3  
(CO1)  
(PO1)

- d) Assuming that RiseTech is relying on cloud computing, which deployment model should they choose? Justify your answer. 5  
(CO1)  
(PO1)
- e) Propose a revenue model for the personalized financial advice provided by the corporation and justify it. 4  
(CO1)  
(PO1)
5. In a world where cancer treatment is a complex landscape, a team of healthcare professionals, data scientists, and app developers comes together to create a mobile app that revolutionizes how patients receive personalized cancer treatments. Patients diagnosed with cancer often face overwhelming treatment options and information overload. Healthcare providers, on the other hand, strive to provide tailored treatment plans using machine learning based on each patient's unique genetic makeup and tumor characteristics. However, patients prefer relying on models that are understandable and interpretable to them when dealing with issues as critical as cancer. Healthcare providers believe that a machine learning model that highlights the treatment plans based on multiple factors considered at different stages of analysis will increase interpretability and hence, reliability.
- Based on the given scenario, answer the following questions:
- a) Which machine learning model would you recommend in such cases? Justify your answer. 6  
(CO3)  
(PO2)
- b) Making necessary assumptions about the user input, design a black-box architectural diagram of the mobile application. 8  
(CO2)  
(PO2)
- c) What are the tools and technologies that you can use to implement the mobile app? 8  
(CO2)  
(PO1)
- d) Mention a legal or ethical issue that might arise in the aforementioned scenario. 3  
(CO2)  
(PO1)
6. a) You and your friends are on the way to try the new ice cream store that uses an automatic ice cream machine. The machine has multiple terminals where you can place the order and a single terminal from where you receive the ice cream. The orders are processed sequentially by a single machine. Your friends wanted to prank you by repeatedly placing incomplete orders to overwhelm the machine. Consequently, the machine was too busy processing the incomplete orders and couldn't serve you. Which cybersecurity attack is analogous to the prank set by your friends? 6  
(CO3)  
(PO2)
- b) Briefly differentiate the following terms: 3 × 4  
(CO1)  
(PO1)
- Database vs. Data Warehouse
  - Machine Learning vs. Traditional Programming
  - Big Picture Thinking vs. Detail-Oriented Thinking
  - Virus vs. Ransomware