



UNIVERSITY OF RUHUNA

Faculty of Engineering

End-Semester 5 Examination in Engineering: March 2026

Module Number: EE5203

Module Name: Machine Learning (C-23)

[Three Hours]

[Answer all questions, all questions carry equal marks]

Write all your descriptions and explanations in point form

Q1 a) Answer using the basic knowledge on Machine Learning (ML).

i) State your own definition for ML by differentiating it with traditional programming. Give an example for its importance in modern data-driven applications.

[1 Mark]

ii) List three different types of ML by giving one algorithm as an example for each.

[1.5 Marks]

b) Answer the question using the knowledge on Decision Trees (DT) Random Forest (RF) models.

i) State the learning approaches, data requirements, and example applications of DT and RF, respectively.

ii) Briefly explain the role of Bootstrap Aggregation (Bagging) in Random Forest.

iii) Describe the process of random sampling with replacement using an example.

[3 x 2 Marks]

c) Briefly Explain the basic concept of Reinforcement Learning (RL) and describe the interaction between the Agent and the Environment with the help of a suitable diagram.

[1.5 Marks]

Q2 Consider the first few instances of a real-world dataset containing customer churn (loyalty) displayed in Table Q1.1. The Churn value 'Yes' means the customer is still with the company and 'No' means the customer has left the company. The dataset is used to predict if a customer will leave the company based on a given set of information.

Table Q2.1

customer ID	Gender	Senior Citizen	Tenure	Total Charges	Contract	Payment Method	Churn
1010	Female	0	12	29	M-to-M	Card	Yes
1011	Male	0	NaN	53.85	M-to-M	Cash	No
1012	Female	0	2	70.7	NaN	Cash	No
1013	Male	1	45	198.65	Annual	Card	No
1014	Female	0	12	29	M-to-M	Card	Yes

- a)
 - i) State the type of Machine Learning problem that you have to address.
 - ii) List the data exploration steps you would follow.
 - iii) List the data cleaning steps you would follow.
 - iv) After you cleaned the data, state one step that you would follow before dividing the data into training and testing.

[4 x 1 Marks]
- b)
 - i) If the number of customers left the company is considerably less than the existing customers, identify the steps you would follow to handle this situation.
 - ii) List the steps that you would include in your pipeline after data is divided into training and testing sets.

[2x 1 Marks]
- c) List two ML models that can be used for the expected prediction.

[1 Mark]
- d) If you are to compare the performance of multiple ML models, list the steps you would follow in order to achieve this task.

[2 Marks]
- e) If you intend to categorize customers based on the given information into different categories, identify the type of ML model you would use to achieve this task.

[1 Mark]

- Q3
- a) Answer the following questions using the knowledge on Deep Learning (DL).
 - i) State the reasons why DL differs from traditional ML techniques.
 - ii) Briefly describe the purpose of activation functions in neural networks and explain two commonly used activation functions.
 - iii) Illustrate the concept of feedforward propagation and explain the role of the loss function in training neural networks.

[3 x 1 Marks]
 - b) Answer Using Knowledge on Convolutional Neural Networks (CNN).
 - i) State the reasons why CNN is suitable for image processing tasks.

[1 Mark]
 - ii) List the main layers in a CNN architecture, Illustrate your explanation with a CNN architecture diagram.

[2 Marks]
 - iii) List any 2 popular CNN architectures.

[1 Mark]
 - c) Answer using the Knowledge on Generative Adversarial Networks (GANs).
 - i) Explain the roles of the Generator and Discriminator networks.

[1 Mark]
 - ii) Explain the training process of GANs and discuss two practical applications of GANs in real-world scenarios

[2 Marks]

- Q4 a) Two ML models (Model A and Model B) are used to perform a binary classification task and the results are given in Table Q4.1 and Table Q4.2, respectively.

Table Q4.1: Results for Model A

		Predicted	
		Positive	Negative
Actual	Positive	50	10
	Negative	5	35

Table Q4.2: Results for Model B

		Predicted	
		Positive	Negative
Actual	Positive	48	12
	Negative	2	38

- i) Calculate the accuracy and the precision for Model A and Model B, respectively. [2 Marks]
- ii) Compare the performance of Model A and Model B using the calculated metrics and determine which model performs better. Justify your answer. [1 Mark]
- b) Suppose that positive and negative in section a) are results from a space lab. A positive result denotes a detecting a threat to the mankind which results in quarantining the subjects and a negative result leads to releasing them to be used by the scientific community.
- i) Calculate the False Positive and False Negative probabilities. [2 Marks]
- ii) Select the best ML model out of the two to be used in this scenario. Justify your answer. [1 Mark]
- iii) Suppose that one of your colleagues argues that none of the models out of the two are suitable for the given scenario, justify the argument. [2 Marks]
- c) List two performance metrics that can be used to compare the performance of ML models other than the ones mentioned in section a) and b). [1 Mark]
- d) Examine the results tables in part a). What data characteristic requires attention during model training? [1 Mark]
- Q5 Suppose that you are assigned a project to detect human activities such as walking, running, cycling etc., using a smart watch. The smartwatch is capable of running an ML model and it consists of a 3-axis accelerometer and a 3-axis gyroscope.
- a) i) Using a block diagram or a flowchart, explain how you would tackle this problem. [2 Marks]
- ii) Which ML problem type would you be solving in this project [1 Mark]

- b) Suppose that the project specifically mentions that an activity should be identified using a series of fixed number of readings from the sensors (accelerometer and the gyroscope). Therefore, each data instance becomes a series of sensor readings and series should overlap by a fixed percentage.
- Identify the ML model you would use and justify your answer.
 - It was identified that there is a data leakage when the data instances are divided into training and testing. Suggest a possible reason for this data leakage.
 - Propose a mechanism to overcome this data leakage.
- [3 x 1 Mark]
- c) Given that the smart watch is to be used by the elderly individuals to track their daily activities, explain your plan for model training. Focus on the quality of samples and ethics.
- [2 Marks]
- d) It is commonly observed that certain activities occur much more frequently than others in daily life. State the method you would use to handle this during model training and evaluation.

[2 Marks]

Equations:

$$E1 = \frac{TP + TN}{TP + TN + FP + FN}$$

$$E2 = \frac{TP}{TP + FP}$$

$$E3 = \frac{FP}{FP + TN}$$

$$E3 = \frac{FN}{TP + FN}$$

$$E5 = \frac{TP}{TP + FN}$$