VIT®

Vellore Institute of Technology
(Deemed to be University under section 3 of UGC Act, 1956)

Continuous Assessment Test II – October-2024

Programme	:	**B. Tech (ECE/ECM)**	Semester	:	**FS 2024-25**
Course	:	**Artificial Intelligence and Machine Learning**	Code	:	**BECE309L**
			Class Nbr	:	**CH2024250100170 CH2024250100200**
Faculty	:	**50441 VIJAYAKUMAR P 53099 NITISH KATAL**	Slot	:	**D1+TD1**
Time	:	**90 Minutes**	Max. Marks	:	**50**

Answer ALL the questions

Q.No.	Sub. Sec.	Questions	Marks	BT Level
1.		An autonomous vehicle must navigate through a city environment, where it encounters two uncertain factors that affect its ability to safely reach its destination: a) *Traffic Density (T):* This can be either high (H) or low (L), and influences the vehicle's speed and decision-making time; and 2nd attribute b) *Weather Conditions (W):* can be clear (C) or rainy (R), which affects road friction and visibility. The probability distributions for Traffic Density (T) and Weather Conditions (W) are:	10	K3

Table 1: Probability distribution for Traffic Density

T (Traffic Density)	Probability
T = High (H)	0.6
T = Low (L)	0.4

Table 2: Probability distribution for Weather Condition

W (Weather Condition)	Probability
W = Clear (C)	0.7
W = Rainy (R)	0.3

The conditional probabilities of the vehicle's Successful Navigation (S) given the combinations of T and W are:

Table 3: Conditional probability distribution for Successful Navigation (S)

T (Traffic Density)	**W (Weather Condition)**	**P(S = Yes)**
T = High (H)	W = Clear (C)	0.8
T = High (H)	W = Rainy (R)	0.3
T = Low (L)	W = Clear (C)	0.9
T = Low (L)	W = Rainy (R)	0.5

Construct a directed a cycling graph-based network to represent this scenario, and calculate the overall probability of the vehicle successfully navigating the environment.

2.		A dataset contains medical records of patients, which includes the following variables: Age, Gender, BMI (Body Mass Index), Blood Pressure (in mmHg), Cholesterol Level (in mg/dL), Smoking Status (Yes/No), Physical Activity Level (Low/Moderate/High), and a target variable, Heart Disease (Yes/No). The dataset contains a total of 10,000 records collected from patients over the last five years. Based upon the above statement answer the following: **[2.5 marks each]**.	10	K2

i. Describe the key data preprocessing steps that you would apply to the dataset before building a machine learning model. Consider handling missing data, data normalization, and dealing with categorical variables (e.g., Gender,

	Smoking Status, and Physical Activity Level). Discuss the importance of each preprocessing step and how it improves the quality of the data for model training. ii. Suggest how domain-specific knowledge could be used to extract new features from the dataset. For instance, how might you derive a risk score from variables like BMI, Blood Pressure, and Cholesterol Level that could serve as a composite indicator of cardiovascular health? iii. Outliers in variables such as BMI or Blood Pressure could distort the model's performance. Suggest the preprocessing methods which can handle outliers in the dataset. What would be preferred; to remove outliers or transform them, and how would you approach affect the overall model accuracy? iv. Assume the dataset contains missing values for BMI and Cholesterol Level. Discuss how they can be addressed?		
3.	A startup is deciding whether to launch a new product, but faces uncertainty about customer demand, production costs, and market competition. Describe the process of simple decision-making under uncertainty in this scenario. How can decision trees be used to help the startup make an optimal decision, considering different possible outcomes? Illustrate your explanation with an example related to product launch success or failure.	10	K4
4.	A financial institution is using customer transaction data to assess credit risk. The dataset consists of the following features for 10,000 customers; the sample data is given as: a) *Transaction History*: Sample data: [120, 80, 200, 150, 220, 0, 300, 90] b) *Credit Score*: Sample data: [720, 650, 780, 610, 800, 540, 500, 730] Due to the large size of the dataset, the institution wants to reduce the dimensionality while retaining the most important features. Use some dimensionality reduction method that uses methods covariance analysis for reducing the features to 1 only.	10	K3
5.	Consider the case study of predictive maintenance system implemented in a manufacturing plant. This system utilizes machine learning algorithms to forecast equipment failures based on sensor data (e.g., temperature, vibration, and operating hours). **[5 Marks Each]** i. Describe the process of analysis of variability of the variation of the certain parameters in the framework of considered predictive maintenance system. How would you determine which input features (e.g., temperature or vibration levels) most significantly impact the model's predictions of equipment failure? ii. Explain how the findings from the such analysis could influence decision-making in the maintenance strategy of the plant. Provide specific examples of potential actions that could be taken based on the sensitivity analysis results.	10	K4

Course Faculty