|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Course Code:** | | | **:** | **AI244AI** | **Date** | **:** | **19-06-2024** | | | |
| **Semester** | | | **:** | **IV** | **Time** | **:** | **12:00 – 1:30 PM** | | | |
| **Max Marks** | | | **:** | **50** | **Duration** | **:** | **90 mins** | | | |
| **Artificial Intelligence and Machine Learning** | | | | | | | | | | |
| **CIE 1** | | | | | | | | | | |
| **Note: Answer all the Questions** | | | | | | | | | | |
| **SL. No** | | **Questions** | | | | | | **M** | **BT** | **CO** |
| **1** | **a** | A chatbot is an artificial intelligence (AI) software that can simulate a conversation (or a chat) with a user in natural language through messaging applications, websites, mobile apps, or through the telephone. Give a PEAS description of a chatbot that is used to promote cultural events in a college and justify your description. | | | | | | 5 | L3 | CO1 |
| **b** | Mention the types of environments suitable for the following tasks and justify your selection briefly;   1. The agent monitors the boiler's temperature in a manufacturing plant. 2. Agent-detecting network attacks | | | | | | 5 | L3 | CO1 |
| **2** | **a** | Differentiate the working of the simple-reflex agents and model-based reflex agents. | | | | | | 5 | L2 | CO1 |
| **b** | By considering any of the domains, such as healthcare, retail, autonomous cars, or others, list four areas for the application of Artificial Intelligence. | | | | | | 5 | L3 | CO1 |
| **3** | **a** | Compare various types of machine learning techniques and their categories with suitable applications. | | | | | | 5 | L2 | CO1 |
| **b** | Illustrate the bias-variance trade-off issue in machine learning. Identify the techniques used to avoid overfitting and underfitting scenarios. | | | | | | 5 | L2 | CO2 |
| **4** |  | Consider the following data: the Y label is whether or not the child goes out to play. Calculate the following:   |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | | **Day** | **Weather** | **Temperature** | **Humidity** | **Wind** | **Play?** | | 1 | Sunny | Hot | High | Weak | No | | 2 | Cloudy | Hot | High | Weak | Yes | | 3 | Sunny | Mild | Normal | Strong | Yes | | 4 | Cloudy | Mild | High | Strong | Yes | | 5 | Rainy | Mild | High | Strong | No | | 6 | Rainy | Cool | Normal | Strong | No | | 7 | Rainy | Mild | High | Weak | Yes | | 8 | Sunny | Hot | High | Strong | No | | 9 | Cloudy | Hot | Normal | Weak | Yes | | 10 | Rainy | Mild | High | Strong | No |  1. Calculate the IG (information gain) for each attribute (feature) 2. Choose which feature to split with and repeat for each level. 3. Choose a feature for each node to split on 4. Build Final Tree | | | | | | 10 | L3 | CO5 |
| **5** | **a** | The following data are recorded while predicting the spam email using a classification model.   1. Correct prediction – 0 spam, 75 non-spam 2. Incorrect prediction- 30 spam, 700 non-spam   Calculate the Accuracy, sensitivity, precision, and F-measure of the model. Identify the problem with the above classification model. | | | | | | 5 | L3 | CO5 |
| **5** | **b** | Analyze and identify issues in the following machine learning scenarios and propose solutions (s) for them   1. A student dataset has an employee age value of 230. 2. An object recognition model training set error is 95%, and the test set error is 15%. 3. Malignant tumour classification model training set error is 25%, and test set error is 23%. | | | | | | 5 | L4 | CO1 |

**M-Marks, BT-Blooms Taxonomy Levels, CO-Course Outcomes**

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Marks Distribution** | **Particulars** | **CO1** | **CO2** | **CO3** | **CO4** | **CO5** | **L1** | **L2** | **L3** | **L4** | **L5** | **L6** |
| **Max Marks CIE** | 30 | 5 | - | - | 15 | - | 15 | 30 | 5 | - | - |

|  |  |
| --- | --- |
| **Course Outcomes: After completing the course, the students will be able to:-** | |
| **CO1** | Explain and apply AI & ML algorithms to address various requirements of real-world problems |
| **CO2** | Design and develop AI and ML solutions to benefit society, science and industry. |
| **CO3** | Use modern tools to create AI and ML solutions. |
| **CO4** | Demonstrate effective communication through team presentations and reports to analyze the impact of AI and ML solutions on society and nature. |
| **CO5** | Conduct Performance evaluation, modelling and validation of AI and ML solutions benefitting lifelong learning. |

Scheme and Solutions

|  |  |  |
| --- | --- | --- |
| Q. No | Solutions | M |
| 1a. | PEAS Description of the Chatbot to promote cultural events in the college;  SAMPLE ANSWER:  Performance: User Engagement, Event Awareness, User Satisfaction, Conversion Rate, Timeliness, etc. (Any TWO – 2M)  Environment: Digital Platforms, User location and demographics, Mobile, etc. (Any 1 – 1M)  Actuators: Notifications, Advertisements, Messages, etc. (Any 1 – 1M)  Sensors: User data, Event data, Data from Social Networks, etc. (Any 1 – 1M) |  |
| 1b. | 1. **The Agent Monitors the Boiler's Temperature in a Manufacturing Plant (2.5M)**   **Suitable Environment: Partially Observable, Dynamic, Continuous, and Deterministic**   1. **Partially Observable:** The agent may not have access to all the environmental variables influencing the boiler's temperature, such as external weather conditions or internal pressure variations, making the environment partially observable. 2. **Dynamic:** The boiler's temperature can change rapidly due to varying operational conditions, heat exchange processes, and potential malfunctions. 3. **Continuous:** The boiler's temperature is a continuous variable that can take on a wide range of values, requiring precise monitoring and control. 4. **Deterministic:** Given the same initial conditions and inputs, the temperature changes predictably based on known physical principles, leading to deterministic behavior.  **Agent Detecting Network Attacks** **Suitable Environment: Partially Observable, Dynamic, Discrete, and Stochastic (2.5M)**   1. **Partially Observable**: The agent may not have complete visibility into all network activities or the internal workings of connected devices, making the environment partially observable. 2. **Dynamic**: Network conditions and potential attack vectors can change rapidly as new devices connect, data traffic fluctuates, and attackers use different techniques. 3. **Discrete**: Network events, such as packet transfers, connection requests, and alerts, are discrete events that occur at specific times. 4. **Stochastic**: The occurrence of network attacks is probabilistic, with uncertainty in both timing and method, resulting in a stochastic environment |  |
| 2a. | Simple Reflex Agent: (Diagram: 1M + Discussion 1.5M )  Topics in AI : AGENTS  Model-based Reflex Agent: (Diagram: 1M + Discussion 1.5M )  Kredo.ai Engineering ... |  |
| 2b. | Each application of AI in a selected domain carries 1.25 M = 1.25 M X 4  EXAMPLE: (Healthcare)   1. To create personalized treatment plans 2. Healthcare data analysis 3. Disease prediction 4. Patient Monitoring |  |
| 3a. | |  |  |  | | --- | --- | --- | | **Supervised machine learning** | **Unsupervised machine learning** | **Reinforcement** | | Used for classifying  (classes and labels are given) | To find pattern  (No idea about class/labels) | Rewards – if guessed correctly  (No idea about class/labels) | | Labelled training data | Unknown and unlabeled data | Model learns and updates itself through rewards/punishment | | Classification and regression | Clustering and association analysis | No such types | | Simple to understand | More difficult to understand and implement | Most complex | | Navie bayes, KNN, LR, SVM etc | Kmeans, PCA, DBSCAN, apriori | Q-Learning, Sarsa | | Hand writing recognition, stock market prediction, disease prediction, fraud detection | Market based analysis, recommender system, customer segmentation | Self driving cars, intelligent robots | | 05 |
| 3b. | Lightbox  To avoid overfitting (high variance), try the following –   * Increase the training data (collecting more data/augment the training dataset) * Less complex model * Early stop (reduce epochs) * Remove features (not recommended) * Assembling * L1/L2 regularization to simplify your model.   To avoid underfitting (high bias):   * Try to increase complexity of model * Train for longer time * Try to increase the number of features by finding new features or making new features from the existing ones. * Decrease regularization * Use different model | 05 |
| 4. | Initial entropy = 𝐻(𝑌) = − ∑ 𝑃(𝑌 = 𝑦) log2 𝑃(𝑌 = 𝑦)𝑦  = −𝑃(𝑌 = 𝑦𝑒𝑠) log2 𝑃(𝑌 = 𝑦𝑒𝑠) − 𝑃(𝑌 = 𝑛𝑜) log2 𝑃(𝑌 = 𝑛𝑜)  = −(0.5) log2(0.5) − (0.5) log2(0.5) = 1    IG = 1-0.7884 = 0.2116        IG(Y, wind) = 0.1245  IG(Y, hum) = 0.1349  IG(Y, weather) = 0.4  IG(Y, temp) = 0.2116 | 10 |
| 5a. | Accuracy=0.8697, sensitivity= 0.0323, precision= 0.0132 and F-measure of the  model. Identify the problem with above classification model. It is imbalanced  dataset. The model failed to predict positive class instances | 05 |
| 5b. | A student dataset has an employee age value as 230.  Outlier, impute with avg, median or other suitable values.  b. An object recognition model training set error is 95% and test set error is  15%. overfitting, decrease model complexity.  c. Malignant tumor classification model training set error is 25% and test set  error is 23%. underfitting, increase model complexity. | 05 |