**Enroll. No.\_\_\_\_\_\_\_\_\_**

**MARWADI UNIVERSITY**

**Faculty of Engineering and Technology**

**Information and Communication Technology**

**B.Tech. SEM: V MU FINAL EXAM December 2023**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Subject: - Machine Learning (01CT1519) Date:- 04/12/2023**

**Total Marks:-50 Time: - 120 minutes**

|  |  |  |
| --- | --- | --- |
| **Question 1: Do as directed: [CO1,CO2] [10]** | | |
| 1. | If you are a machine learning engineer working in the industry, how will you select supervised  algorithm. Also give the reason of selection. | [02] |
| 2. | Supervised algorithm has training and testing phases. Is there any algorithm which does not have training phase? Justify your answer. | [02] |
| 3. | What do you mean by outliers ? Which of the unsupervised learning algorithm works well in the present of outliers ? How ? | [02] |
| 4. | Demonstrate non linear boundary of EX-OR gate. Show EX-OR and AND gate realization with help of neural network and weights. | [02] |
| 5. | Draw the box plot for the given statistics of the dataset. Comment on the nature of the features, based on the box-plot obtained. | [02] |
| **Question 2: Solve the following [Attempt any two]: [18]** | | |
| 1. | For the given set of points identify clusters using single linkage and draw the Dendrogram. How many clusters are good to consider in this case ? | [09] |
| 2. | For the sunburn data set below construct decision tree.   |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | | **Name** | **Hair** | **Height** | **Weight** | **Location** | **Class** | | Avni | Blonde | Average | Light | No | Yes | | Ramesh | Blonde | Tall | Average | Yes | No | | Shailja | Brown | Short | Average | Yes | No | | Khushali | Blonde | Short | Average | No | Yes | | Amarpreetsingh | Red | Average | Heavy | No | Yes | | John | Brown | Tall | Heavy | No | No | | Yasin | Brown | Average | Heavy | No | No | | Namrata | Blonde | Short | Light | Yes | No | | [09] |
| 3. | Show by taking the dummy example with (not less than) 2 features, the requirement of the normalization of data, in predicting the output value. Explain how the model may behave adversely, in case the normalization is not involved in the pre-processing step. **[CO1, CO3]** | [09] |
| **Question 3: Do as directed [08]** | | |
| 1. | For the given confusion matrix, calculate various performance parameters and comment on it. **[CO4]**   |  |  | | --- | --- | | 1042 | 12 | | 35 | 43 |   **True label**  **Predicted Label** | [02] |
| 2. | Calculate various performance metrics for the linear regression. **[CO4]**   |  |  |  |  | | --- | --- | --- | --- | | **Sr. No.** | **Xi**  **(size of house)** | **Yi**  **Actual (in Lac)** | **Y (hat)**  **Estimated**  **(in Lac)** | | 1 | 80 | 40 | 42 | | 2 | 90 | 46 | 45 | | 3 | 120 | 72 | 70 | | 4 | 105 | 64 | 67 | | 5 | 72 | 35 | 34 | | 6 | 145 | 112 | 115 | | [02] |
| 3. | Explain the process of using “Random Forest” as   1. Classification Problem 2. Regression Problem   How would you evaluate the designed ML model for both the cases? **[CO1, CO4]** | [02] |
| 4. | You are designing a deep learning system to detect driver fatigue in cars. It is crucial that your model detects fatigue, to prevent any accidents. Design the system that achieves the given problem statement. Also state, which of the following is the most appropriate evaluation metric : Accuracy, Precision, Recall, Loss Value. Explain your choice **[CO4]** | [02] |
| **Question 4: Do as directed [Attempt any 3]: [09]** | | |
| 1. | After visually inspecting the dataset for classifying the historical places, you realize that the training set only contains pictures taken during the day, whereas the validation set only has pictures taken at night. Explain what is the issue and how you would correct it. **[CO5]** | [03] |
| 2. | Suppose you have to rotate an image. Image rotation is nothing but multiplication of image by a specific matrix to get a new transformed image.    consider one point in the image to rotate with co-ordinates as (1, 0) to a co-ordinate of (0, 1), identify the kernel matrix for the given task? **[CO5]** | [03] |
| 3. | For any given CNN model, the accuracy vs epochs and loss vs epochs graphs are as given below.    Interpret the graph obtained after training the model. Is the trained model, optimized? If not, what possible changes can be done in model architecture to improve the accuracy and loss graphs? Draw the improved accuracy and loss graphs. **[CO5]** | [03] |
| 4. | Consider the summary of the designed sequential model. Draw the architecture of the mentioned sequential model with all the necessary model configurations. **[CO5]** | [03] |
| **Question 5: Do as directed [Attempt any 1]: [05]** | | |
| 1. | Explain the training, validation and testing dataset and their importance in designing any ML model. Explain the significance of regularization w.r.t. bias and variance observed for any designed ML model considering overfitting and underfitting concepts. **[CO3]** | [05] |
| 2. | Explain the process of feature selection and process to remove the redundant features from the dataset. Why the redundant features are required to be removed for training the model. Explain giving examples. [**CO3]** | [05] |

**---Best of Luck---**