SHRI MADHWA VADIRAJA INSTITUTE OF TECHNOLOGY & MANAGEMENT, BANTAKAL

Department: Artificial intelligence & data science	Assignment- I	Academic Year: 2023-24
Class: 6th semester	Course: Machine learning	Course Code: 21AI63
Date: 08-07-2024	Submission date : 19-07-2024	Max. Marks: 10

Qn. No	Question	Marks	TLO*	BL*	CO*
1	Consider the Titanic dataset. The goal is to train classifiers that can predict the Survived column based on the other columns. Implement the following tasks one by one in Python: 1. Load the dataset and display the initial rows in the dataset. 2. Use the describe() function to generate a summary of the data in the dataset. 3. Determine whether any of the columns in the dataset have missing values. 4. Filter the dataset to remove columns that won't be used, one-hot-encode the "Sex" and "Pclass" columns and remove rows with missing values. 5. Split the data into two datasets: one for training and one for testing. Use a stratified split to create a balanced distribution of samples in the training dataset and the testing dataset. 6. Build and train any three machine learning models on the training dataset. 7. Compute the classification accuracy for the three classifiers for the test set. 8. Display the corresponding confusion matrices. 9. Print a classification report to view precision, recall, and F1-score. 10.Plot the ROC curves for the three classifiers in a single plot and display the AUC values. **Answer the following questions:** 1. Describe in detail the attributes of the dataset. 2. Which are the attributes that have missing values? 3. What is the train: test split ratio considered? How many samples are in the training and test sets? 4. What are the three machine learning models that you have considered? Mention the values for the hyperparameters of the model. 5. Which model has given the highest accuracy? 6. Discuss the performance of the three classifiers (in terms of correct and incorrect classifications) with the help of the confusion matrices. 7. What can you infer from the ROC curves?	10	2.1,2.2	L3	CO2

BL* Bloom's Taxonomy Level; CO* Course Outcome; TLO – Topic Level Outcomes