Department of Physical Science
Faculty of Applied Science
In-course Assessment Examination – ICAE 01
Machine Learning (P)-CSH4144(P)

Time: 45 minutes **Date: - 17.12.2024**

01. The dataset comprises information collected in a restaurant setting to analyze factors influencing **tip amounts**. The goal is to predict the tip amount based on several explanatory variables. Each row in the dataset represents a single observation of a dining experience. The features include Total bill, Tip amount, sex, smoker, time, day, and size. Build a model using Python programming language by completing the following steps:

    a. Load the data into a dataset from the file tips_dataset.csv. **[05 marks]**

    b. Display the last 5 rows of the dataset. **[05 marks]**

    c. Encode the original data into numerical data for the following columns as prescribed:

        i. Sex: Male: 0, Female: 1 **[05 marks]**

        ii. Smoker: No: 0, Yes: 1 **[05 marks]**

        iii. Time: Lunch: 1, Dinner:2 **[05 marks]**

        iv. Day: Mon: 1, Tue: 2, Wed: 3, Thur:4, Fri:5 **[05 marks]**

    d. Select the columns, as features that are affecting the amount of the tip: (Sex does not much affect the amount of the tip). **[05 marks]**

    e. Split the dataset into a training and testing set such that 30% of the data is for the test set. **[05 marks]**

    f. Fit a multiple linear regression model by using the training set split before. **[10 marks]**

    g. After training your model, predict the values for test data. **[05 marks]**

    h. Display the intercept and the coefficient. **[10 marks]**

    i. Evaluate the model's performance, using:

        i. R Squared **[05 marks]**

        ii. Mean Squared Error (MSE) **[05 marks]**

        iii. Mean absolute error (MAE) **[05 marks]**

        iv. Root mean square Error (RMS) **[05 marks]**

    j. Compute and display the actual values and its predicted values for the first five records. **[05 marks]**

    k. Draw a straight line to fit the linear regression model of the given dataset. **[10 marks]**