**Heaven’s Light is Our Guide**



*Department of Computer Science & Engineering*

*Rajshahi University of Engineering & Technology*

Lab Manual

**Course Title:** Sessional based on CSE 4119

**Course No.**: CSE 4120

**Credit:** 0.75

**Reference Book:**

[1] Jiawei Han, Micheline Kamber, and Jian Pei, “Data Mining: Concepts and Techniques”

**Experiments/Problems:**

**[Week 1]** Introduction to Data Mining and Tools

**Objective:** Understand basic data mining concepts and tools.

**Topics:** Data types, tasks (classification, clustering, association).

**Tools:** Weka, Python, or Orange.

**Datasets:**

* *Iris Dataset* – Classification, clustering
* *Online Retail* – Association rule mining (market basket) and Clustering
* *House Prices* – Regression

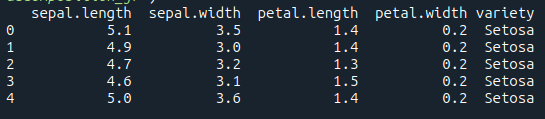
**Tasks:**

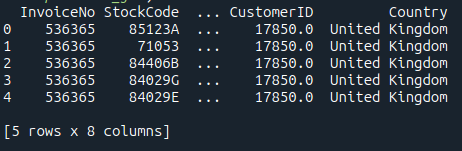
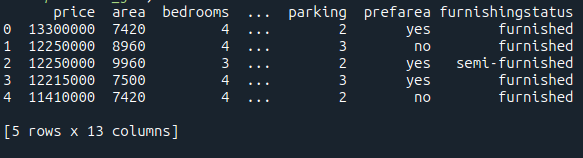
* Install Python libraries (e.g., Pandas, scikit-learn, Matplotlib).
* Explore datasets using pandas.

**Lab Outcomes:**

* **Load dataset (in .csv or .xlsx format):**

1. *Iris Dataset*



1. *Online Retail*
2. *House Prices*

* Try to understand the data type. Determine the dimensions of the data.
* Find out which one you can use as a label.
* Try different ways to split the data into training, validation, and test sets (60% training, 20% validation, and 20% test data).
* Identifying the dependent and independent variables.
* Split into two sets: features and labels.
* Identifying the number of labels and showing them. Plot them using Matplotlib.
* Based on the labels categorised, what type of problem is this: classification, multiclass classification, regression or clustering?