**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 2]** Data Preprocessing

**Objective:** Apply preprocessing techniques to raw data.

**Topics:** Missing values, normalization, binning, encoding.

**Tasks:**

* Load a dataset and handle missing values.
* Dealing with categorical data: Label encoding and one-hot encoding of categorical variables
* Scaling the features: Perform min-max normalisation and Z-score standardisation.
* Similarity and Dissimilarity Measures:
  1. Pearson’s Correlation
  2. Cosine Similarity
  3. Jaccard Similarity
  4. Euclidean Distance