**BIN2023R01 – INTRODUCTION TO DATAMINING & MACHINE LEARNING FOR BIOINFORMATICS**

Lab Exercise 4- Normalization, standardization, and feature transformation techniques (incl. discretization)

Aim: To perform normalization, standardization, feature transformation, and discretization on the iris dataset

Procedure:

**1. Load iris dataset.**

**2. Perform normalization using the Max-Min function and interpret the results before and after normalization.**

**3. Perform standardization using z-score normalization and interpret the results before and after standardization.**

**4. Perform feature transformation by applying function transform (Log transform and square root transform) and power transform (Box-Cox transform). Interpret the results before and after the feature transformation.**

**5. Perform discretization by applying equal-width binning and equal-depth binning. Interpret the results before and after discretization.**

**Questions:**

1. What is the purpose of normalization in machine learning, and how does it differ from standardization?

2. In what scenarios would you prefer normalization over standardization, and vice versa?

3. What are the potential advantages and disadvantages of applying feature transformation techniques to a dataset?

4. How would you interpret the results of feature transformation techniques visually?

5. Describe the process of discretization and its significance in building the machine learning model. What are some common discretization techniques?

5. What is the difference between equal-width binning and equal-depth binning? Interpret the results of each binning method of iris datasets.

6. How does the log transform differ from the square root transform? Which transform can be used for which type of data?

7. What is function transform and power transform? When might you choose one type of transformation over the other?

**Soft copy deadline: ------ 11:59PM**

**Hard copy deadline: ------ 3:15PM**