School of Computer Science and Artificial Intelligence, SR University

Course Code	Course Title	L	R	P	Cr
25CAI202PC501	Advanced Data Science	2	0	2	3
Program Core					

MACHINE LEARNING LAB TEST

Time : 2 hours

Total marks : 20

Dr. Sumit Kumar Singh

Associate Professor

Question 1 (10 Marks)

Using the dataset **"KNN_regression_dataset.xlsx"**, perform the following tasks:

1. Load the dataset and display summary statistics of all features.

2. Split the data into **80% training** and **20% testing** sets.

3. Standardize/scale the input features appropriately.

4. Build a **KNN Regression model with k = 5** to predict Efficiency (%).

5. Use the model to predict the target values for the test set.

6. Calculate and report the following evaluation metrics:

- Mean Absolute Error (MAE)
- Root Mean Squared Error (RMSE)
- R^2 Score

Question 2 (10 Marks)

Study the effect of different **k-values** on the performance of KNN Regression using the same dataset.

1. Train four separate KNN Regression models with:
k = 1, 3, 5 and 10

2. For each model, compute and record the **R^2 Score** on the test data.

3. Plot a graph of **k vs R^2 Score**.

4. Identify which value of k gives the best performance.

Marking Scheme 1 (10 marks):

- Data loading & summary → **2 marks**
- Splitting & scaling → **2 marks**
- Training KNN model → **2 marks**
- Evaluation metrics → **4 marks**

Marking Scheme 2 (10 marks):

- Models for all k values → **3 marks**
- R^2 calculation → **3 marks**
- Plot (k vs R^2) → **2 marks**
- Best k → **2 marks**