Course Code: CS481   **Course Name:**   **Introduction to Data Science**

STRICTLY NO MOBILE PHONES ARE ALLOWED IN THE EXAMINATION

Time allowed: 1 hour   **Student Roll #_____**
Maximum Marks: **Please return complete Test paper.**

ATTEMPT ALL QUESTIONS ON THE ANSWER SHEET. DON'T WRITE ANYTHING ON THE QUESTION PAPER

Instructions:

- **You must sign and print your name on the honor statement** and return the question paper.
- Read each question completely before answering it. There are **3 questions** and **1 page**.
- In case of any ambiguity, you may make assumption. But your assumption should not contradict any statement in the question paper.
- All the answers must be solved according to the sequence given in the question paper.

Time: 60 minutes.   **Max Marks**: 10 points

Honor Statement:

I solemnly affirm that I will not (and have not) copy or cheat during the exam.

---   ----------------------------

(Roll Number)   (Signature)

Question 1: Multiple Choice Questions (Removed)

Question 2: [3.5 Points]

As a data scientist, you got a project from the Traffic Police to know the factors of speeding in certain areas. What are the main steps you will do to accomplish the above mentioned task?

Question 3: [4 Points]

Use Leave one out Cross Validation and Nearest Neighbour classifier to find the predicted class for each sample in Table below. Calculate accuracy, precision, recall, and F-measures of the model. There are 4 instances and 2 attributes (Att1, Att2). Use cityblock distance: $d(p,q) = \sum_i |p_i - q_i|.$ See Appendix for formulas. Show all steps clearly

Instance #	Att1	Att2	Actual Class
1	2	3	0
2	1	5	1
3	4	2	1
4	2	5	0
5	6	8	0

Appendix

$$Accuracy = \frac{TN + TP}{TN + FN + TP + FP}$$

$$recall = \frac{TP}{TP + FN}$$

$$precision = \frac{TP}{TP + FP}$$

$$F_1 = 2 \cdot \frac{precision \cdot recall}{precision + recall}$$

BEST OF LUCK!