



SPRING MID SEMESTER EXAMINATION-2024

School of Computer Engineering
Kalinga Institute of Industrial Technology, Deemed to be University
Data Mining and Data Warehousing
[IT-3031]

Time: 1 1/2 Hours

Full Mark: 20

*Answer Any four questions including question No.1 which is compulsory.
The figures in the margin indicate full marks. Candidates are required to give their answers in their own words
as far as practicable and all parts of a question should be answered at one place only.*

1. Answer all the questions. [1 Mark X 5]
 - a) Define Data mining and discuss its importance in Information Technology.
 - b) Explain min-max normalization with a suitable example.
 - c) What is the difference between probabilistic and non-probabilistic sampling?
 - d) What is Jaccard Coefficient? Explain with one example.
 - e) Explain the term ETL with respect to data warehouse.
2.
 - a) Explain Data Mining as a step in KDD with a suitable figure. [2.5 Marks]
 - b) Explain with suitable examples different kinds of Attributes in data mining. [2.5 Marks]
3. Find the Cosine distance matrix & Manhattan distance matrix of below points. [5 Marks]

Point	X	Y
P1	0	2
P2	5	0
P3	4	3
P4	6	2

4.
 - a) Define Support and Confidence in the context of frequent pattern mining. [2.5 Marks]
 - b) What is Apriori property? Using Apriori algorithm find all the association rules with minimum support 50 % and confidence 30% for the following dataset. [2.5 Marks]

TID	Data Items
101	Book, Note, Pen
102	Pencil, Note, Eraser
103	Book, Pencil, Note, Eraser
104	Pencil, Eraser, Book
105	Note, Pencil

5. a) What is Bessel's correction? Calculate and draw a Box plot for the dataset 5, 5, 8, 9, 6, 3, 3, 12, 2, 6, 2, 7, 3, 9, 11 [2.5 Marks]
- b) The ages of 112 people lives in a island are given in the following dataset. Find the estimated mean, estimated median and estimated mode of the data set. [2.5 Marks]

Age	Count
0-9	20
10-19	21
20-29	23
30-39	16
40-49	11
50-59	10
60-69	7
70-79	3
80-89	1

*** Best of Luck ***