Code No: 157BC **R18**

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY HYDERABAD
B. Tech IV Year I Semester Examinations, February - 2025
DATA MINING
(Common to CSE, IT, ITE)

Time: 3 Hours **Max. Marks: 75**

Note: i) Question paper consists of Part A, Part B.
ii) Part A is compulsory, which carries 25 marks. In Part A, Answer all questions.
iii) In Part B, Answer any one question from each unit. Each question carries 10 marks and may have a, b as sub questions.

PART – A

(25 Marks)

1.a)	Define Data Integration.	[2]
b)	Define Data mining task primitives.	[3]
c)	Define Confidence and support.	[2]
d)	Discuss support and confidence formulas.	[3]
e)	Define Gini Index.	[2]
f)	Define types of Lazy Learners.	[3]
g)	Define Inter Cluster similarity.	[2]
h)	What is Grid based clustering technique?	[3]
i)	What are Transactional databases?	[2]
j)	Define applications of spatial data mining.	[3]

PART – B

(50 Marks)

2.a) What is Data Preprocessing and explain various techniques.
b) Discuss the various methods for data normalization. [5+5]

OR

3. Explain the architecture of Data warehouse with a neat diagram. [10]

4. What is an Association Rule Mining? Give an Illustration and explanation of Apriori algorithm. [10]

TID	Items
100	F, A, C, D, G, I, M, P
200	A, B, C, F, L, M, O
300	B, F, H, J, O, W
400	B, C, K, S, P
500	A, F, C, E, L, P, M, N

OR

5.a) Discuss correlation analysis.

b) Discuss about mining multilevel and quantitative association rules from transaction databases. [5+5]

6. How does the Naive Bayesian classification works? Explain. How to measure classifier accuracy? [10]

OR

7. Explain decision tree learning algorithm. Apply this algorithm on the following dataset to generate classification rules. Loan application dataset: [10]

Id	Age	Has-job	Own-house	Credit-rating	Class
1	Young	False	False	Fair	No
2	Young	False	False	Good	No
3	Young	True	False	Good	Yes
4	Young	True	True	Fair	Yes
5	Young	False	False	Fair	No
6	Middle	False	False	Fair	No
7	Middle	False	False	Good	No
8	Middle	True	True	Good	Yes
9	Middle	False	True	Excellent	Yes
10	Middle	False	True	Excellent	Yes
11	Old	False	True	Excellent	Yes
12	Old	False	True	Good	Yes
13	Old	True	False	Good	Yes
14	Old	True	False	Excellent	Yes
15	Old	False	False	Fair	No

8.a) Define clustering. Mention the types of clustering techniques.

b) Explain OPTICS clustering with DBSCAN. What are the similarities and differences? [5+5]

OR

9.a) Explain about k-nearest neighbor clustering technique.

b) Differentiate AGNES with DIANA clustering methods. [5+5]

10. Explain

a) Mining the Textual database.

b) Mining Time – Series data. [5+5]

OR

11. Illustrate

a) Mining the World Wide Web.

b) Spatial data mining. [5+5]

---ooOoo---