



Register Number:

Date:

ST. JOSEPH'S COLLEGE (AUTONOMOUS), BANGALORE-27
M.Sc BIG DATA ANALYTICS - II SEMESTER
SEMESTER EXAMINATION: APRIL 2019
BDA 2418- DATA MINING

Time- 2 1/2 hrs

Max Marks-70

This paper contains one printed pages and one parts

Answer any Seven questions of the following.

All question carries equal marks (10 marks)

1. What is data mining? Explain the concepts in data mining using a block diagram.
2. A chain is only Explain about data ware house and different types of data.
3. Why is data visualization important? Explain different types of data visualization used in industry.
4. Write the steps involved in pre-processing of data.
5. What is similarity? Explain different types of similarity and find similarity for the following Phrases
 - a. as strong as its weakest link
 - b. A countenance more in sorrow than in anger
 - c. A Daniel come to judgement
6. Solve the problem using clustering algorithm.

x	2.5	0.5	2.2	1.9	3.1	2.3	2	1	1.5	1.1
y	2.4	0.7	2.9	2.2	3	2.7	1.6	1.1	1.6	0.9

7. Why is recommendation engine an integral part of modern business? Explain with appropriate example.
8. Explain Information gain. Explain CART algorithm for the given data.

Day	Outlook	Temp.	Humidity	Wind	Decision
1	Sunny	Hot	High	Weak	No
2	Sunny	Hot	High	Strong	No
3	Overcast	Hot	High	Weak	Yes
4	Rain	Mild	High	Weak	Yes
5	Rain	Cool	Normal	Weak	Yes

9. Explain Naïve Bayesian classification with appropriate example.