



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. Explain the process of knowledge discovery with help of a Block Diagram.
2. What is the difference between structured data, semi structured and unstructured data? Explain the different attributes of data.
3. "*Business Intelligence is vital tool for any business success*" Do you agree or disagree the given statement, if so, with help of a case study defend your argument.
4. Explain pre-processing of data in exact 1000 words.
5. What is similarity? Explain different types of similarity and find similarity for the following Phrases
  - Julie loves me more than Linda loves me
  - Jane likes me more than Julie loves me
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. Define Entropy with example. Explain ID3 algorithm for the given data.

Item Identifier	Item Weight	Item Fat Content	Item Type	Item Purchased
FDA15	9.3	Low Fat	Dairy	1
DRC01	5.92	Regular	Soft Drinks	1
FDN15	17.5	Low Fat	Meat	0
FDX07	19.2	Regular	Fruits and Vegetables	1
NCD19	8.93	Low Fat	Household	1
FDP36	10.395	Regular	Baking Goods	0

9. Explain Naïve Bayesian classification with appropriate example.