15. (a) What are the major differences between searching conventional text and searching the web?

Or

(b) Write about web content mining.

PART C — (5 × 8 = 40 marks)

Answer ALL questions choosing either (a) or (b).

Each answer should not exceed 600 words.

16. (a) Explain the Data mining process in details.

Or

(b) Describe the different data mining Techniques.

17. (a) Explain Apriori Algorithm.

Or

(b) Explain FP–Growth with an example.

18. (a) Explain SPLIT algorithm based on Information theory.

Or

(b) Explain Naive Bayes method.

19. (a) Explain K–Means method for clustering.

Or

(b) Describe the methods for computing distance between clusters.

20. (a) Write about web usage mining.

Or

(b) Explain web structure mining.

———————

B.Sc. (CBCS) DEGREE EXAMINATION, APRIL 2019.

Sixth Semester

Computer Science/ Software Engineering — Main

DATA MINING

(For those who joined in July 2016 onwards)

Time : Three hours   Maximum : 75 marks

PART A — (10 × 1 = 10 marks)

Answer ALL questions.

Choose the correct answer :

1. A warehouse is designed to have summary information collected from one or more ————.

(a) OLTP systems

(b) Data bases

(c) Organizations

(d) Decision making tools

2. Which step is used to find the best source database for the data that is required
   (a) Requirement Analysis
   (b) Data selection and collections
   (c) Cleaning and preparing data
   (d) Validation

3. Given the Association rules $AB \to c$ and $xy \to z$ which one of the following is true?
   (a) $Z \to A$
   (b) $y \to xy$
   (c) $A \to C$
   (d) $B \to A$

4. Support (xy) = ?
   (a) $p(x)$
   (b) $p(y)$
   (c) $p(x \eta y)$
   (d) $p(y/x)$

5. Decision tree is a model that is ————
   (a) Predictive
   (b) Descriptive
   (c) Both (a) and (b)
   (d) None

6. Classification is ————
   (a) Separation of objects in to classes
   (b) Ordering of thing in to classes
   (c) Separation of things in to classes
   (d) All the above

7. The desired feature of cluster Analysis is ————
   (a) More scan of the dataset
   (b) Maximum input parameters
   (c) Ability to discover more cluster shapes
   (d) Not scalable

8. In cluster analysis Distance ————
   (a) is always positive
   (b) from point x to itself is always zero
   (c) from x to y is always the same as from y to x
   (d) all the above

9. ———— deals with understanding user behaviour is interacting with the web or with a web site.
   (a) web content mining
   (b) web structure mining
   (c) web usage mining
   (d) hyper link

10. Which algorithm is used for web structure mining?
    (a) DIPRE
    (b) STC
    (c) HITS
    (d) Finger printing

PART B — (5 × 5 = 25 marks)

Answer ALL questions choosing either (a) or (b).

Each answer should not exceed 250 words.

11. (a) What is data mining? Describe any four applications of related groups.

    Or

    (b) List the data mining software and give its importance.

12. (a) Describe Apriori – TID with an example.

    Or

    (b) Write about DHP.

13. (a) Explain the concept of classification.

    Or

    (b) Describe the TREE induction algorithm.

14. (a) Describe the features of cluster Analysis.

    Or

    (b) How should clustering be carried out for very large data sets?