

Reg No.: _____

Name: _____

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY
FIFTH SEMESTER REGULAR AND THIRD SEMESTER SECOND YEAR DIRECT MCA
DEGREE EXAMINATION(S) MAY 2019
Course Code: RLMCA301
Course Name: WEB DATA MINING

Max. Marks: 60

Duration: 3 Hours

PART A*Answer all questions, each carries 3 marks.*

Marks

- | | | |
|---|--|-----|
| 1 | Define World Wide Web. List any two unique characteristics of Web. | (3) |
| 2 | What is Entropy in decision tree learning? | (3) |
| 3 | Explain the concept of Precision and Recall | (3) |
| 4 | Write the k-means algorithm. | (3) |
| 5 | Explain any three types of user queries with example. | (3) |
| 6 | Write a note on any two integer compression methods. | (3) |
| 7 | Explain any two Hiding Techniques used by the spammers? | (3) |
| 8 | What is Web Usage Mining? Mention the goal of Web Usage Mining | (3) |

PART B*Answer six questions, one full question from each module and carries 6 marks.***Module I**

- | | | |
|---|---|-----|
| 9 | Write the Apriori algorithm for generating frequent itemsets. | (6) |
|---|---|-----|

OR

- | | | |
|----|--|-----|
| 10 | a) Write the algorithm for mining sequential pattern based on GSP. | (4) |
| | b) Explain the two steps of Candidate-gen-SPM | (2) |

Module II

- | | | |
|----|---|-----|
| 11 | What is SVM? Explain Linear SVM Non separable case. | (6) |
|----|---|-----|

OR

- | | | |
|----|--|-----|
| 12 | What is Hierarchical Clustering? Write the Agglomerative Hierarchical Clustering algorithm | (6) |
|----|--|-----|

Module III

- | | | |
|----|--|-----|
| 13 | Write a note on | |
| | a) Boolean Model | (3) |
| | b) TF-IDF Scheme of Vector Space model | (3) |

OR

- 14 What is IR System? Explain the general IR system architecture with a neat diagram. (6)

Module IV

- 15 What is Duplicate detection? Describe n - grams duplicate detection technique (6)

OR

- 16 What is Latent Semantic Indexing? Describe the technique used to estimate the latent structure. (6)

Module V

- 17 What is spamming? Explain in detail various Link Spamming techniques (6)

OR

- 18 Write a note on the following Implementation issues of web crawling
- a) Link Extraction and Canonicalization (2)
 - b) Spider Traps (2)
 - c) Concurrency (2)

Module VI

- 19 Describe the following task in Web Usage Data Pre-Processing
- a) Sessionization (2)
 - b) Path Completion (2)
 - c) Data Integration (2)

OR

- 20 Explain any three types of pattern discovery and analysis techniques employed in the Web usage mining. (6)
