

Reg No.: _____

Name: _____

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY
THIRD SEMESTER MCA (LE) DEGREE EXAMINATION, DECEMBER 2017

Course Code: RLMCA 301

Course Name: WEB DATA MINING

Max. Marks: 60

Duration: 3 Hours

PART A

Answer all questions, each carries 3 marks.

		Marks
1	What is Web Mining? What are the Web Mining task?	(3)
2	How can you measure the amount of impurity in decision tree learning?	(3)
3	What is F score? Explain its importance.	(3)
4	What is the importance of Linear SVM: Non-separable Case?	(3)
5	How are the documents represented in Boolean Model?	(3)
6	What is Stemming?	(3)
7	Write a note on Breadth-First Crawlers.	(3)
8	Which are the four primary groups of data obtained from various sources in Web usage mining?	(3)

PART B

Answer six questions, one full question from each module and carries 6 marks.

Module I

- 9 a) Write the Apriori algorithm for generating frequent item sets. (4)
 b) Explain the steps of Candidate-gen function. (2)

OR

- 10 Explain the unique characteristics and challenges of Web. (6)

Module II

- 11 a) Write the Decision Tree Learning Algorithm. (4)
 b) Explain the concept of information gain (2)

OR

- 12 a) Explain K-means Algorithm. (3)
 b) What are the strengths and weaknesses of K-means? (3)

Module III

- 13 Explain Vector Space Model. (6)

OR

- 14 a) Explain the general IR System architecture. (3)
 b) What are the different forms of user queries? (3)

Module IV

- 15 a) Explain the concept of Inverted Index. (2)
 b) What are the steps for searching relevant document in the Inverted Index? (4)

OR

- 16 a) Explain SVD? (4)

A

A7405

- b) What is the important feature of SVD? (2)

Module V

- 17 How is Out-Link Spamming different from In-Link Spamming? (6)

OR

- 18 Draw and explain the flow chart of basic sequential crawler. (6)

Module VI

- 19 Explain the essential task in Web Usage Data Pre-Processing. (6)

OR

- 20 Explain the concept of Data Modelling for Web Usage Mining. (6)
