

Reg. No. :

--	--	--	--	--	--	--	--	--	--

Question Paper Code: 42938

M.E.DEGREE EXAMINATION, NOV 2016

Elective

Computer Science and Engineering

14PCS508 - WEB DATAMINING

(Regulation 2014)

Duration: Threehours

Maximum: 100 Marks

Answer ALL Questions

PART A - (5 x 1 = 5 Marks)

- Name the rules that cannot be found, if the minimum support is set too high in the data set.
(a) frequent rules (b) rare rules (c) infrequent rules (d) both b and c
- K-nearest neighbor is one of the
(a) Learning technique (b) OLAP tool
(c) Purest search technique (d) Data warehousing tool
- Odd one out?
(a) Popularity (b) Authority (c) Prestige (d) Impact factor
- The extraction for an webpage is done using a tree structure
(a) DOM Trees (b) FP tree (c) EC tree (d) None of the above
- Which of the following is not a part of the web log preparation process?
(a) De-spidering (b) Sessionization (c) Path completion (d) Reporting

PART - B (5 x 3 = 15 Marks)

6. What is the use of meta search?
7. Differentiate supervised learning and unsupervised learning.
8. Compare co-citation and bibliographic coupling.
9. List out shortcomings of wrapper generation using supervised learning.
10. What you meant by pre-processing of web usage data?

PART - C (5 x 16 = 80 Marks)

11. (a) Express the methods involved in generating rules from sequential patterns in detail. (16)

Or

- (b) Write any two algorithm for association rule mining with examples. (16)

12. (a) Apply the process of learning from labeled and unlabeled examples in detail. (16)

Or

- (b) Discuss in detail about unsupervised learning models. (16)

13. (a) Summarize HITS algorithm and page rank algorithm and compare its features in detail. (16)

Or

- (b) Describe the working of web community discovery. (16)

14. (a) How to extracting evolution of web community from a series of web archive. (16)

Or

- (b) Compare string matching and tree matching with examples in detail. (16)

15. (a) Explain in details about web usage mining using probabilistic latent semantic analysis. (16)

Or

- (b) Discuss in detail Latent Dirichlet allocation model. (16)