

Reg No.									
----------------	--	--	--	--	--	--	--	--	--



MANIPAL INSTITUTE OF TECHNOLOGY
 (Constituent Institute of Manipal University)
 MANIPAL-576104



VII SEMESTER B.E. (CSE) DEGREE END SEMESTER EXAMINATION
DECEMBER 2010
DATA WAREHOUSE AND DATA MINING (CSE- 405.3)
(ELECTIVE)
- 12 -2010

TIME: 3 HOURS

MAX.MARKS: 50

Instructions to Candidates

- Answer **any five** full questions.
- Missing data can be suitably assumed.

- 1 A) Give a formal definition of a data warehouse and explain the important characteristics of a data warehouse with examples. 4M
- B) What are concept hierarchies and how are they useful in OLAP ? 4M
- C) What is METADATA? Briefly discuss. 2M
- 2 A) Explain with the aid of neat diagram the Knowledge discovery in databases (KDD). 4M
- B) Elaborate on the usage of data mining technique which suits for Market basket analysis. 2M
- C) Write an algorithm for construction of FP –Tree and Explain. 4M
- 3 A) Write a partition based clustering algorithm which handles outliers efficiently. 4M
- B) List any four issues and challenges in Data Mining. 4M
- C) Distinguish between clustering and classification techniques with respect to data mining. 2M

- 4 A) What is downward closure property? Explain with an example. 2M
 B) Write A priori algorithm used to find frequent item sets. 4M
 C) State Bayes theorem. Also explain how Bayes theorem can be used for classification in Data Mining. 4M
- 5 A) For the table given below, determine the best split attribute for first level in constructing a Decision tree using information gain (entropy) approach. 5M

<i>Outlook</i>	<i>Temperature</i>	<i>Humidity</i>	<i>Windy</i>	<i>Play?</i>
sunny	hot	high	false	no
sunny	hot	high	true	no
overcast	hot	high	false	yes
rainy	mild	high	false	yes
rainy	cool	normal	false	yes
rainy	cool	normal	true	no
overcast	cool	normal	true	yes
sunny	mild	high	false	no
sunny	cool	normal	false	yes
rainy	mild	normal	false	yes
sunny	mild	normal	true	yes
overcast	mild	high	true	yes
overcast	hot	normal	false	yes
rainy	mild	high	true	no

- B) Define CF tree. What are its advantages? 2M
 C) Discuss any two types of variable used in clustering with examples. 3M
- 6 A) What is PageRank? How is it computed? 2M
 B) What is web mining? Explain the different categories of web mining. 4M
 C) What is learning? Explain Back Propagation learning algorithm used in neural networks. 4M

