UNIVERSITY EXAMINATIONS: 2013/2014

EXAMINATION FOR THE MASTER OF SCIENCE IN INFORMATION

SYSTEMS MANAGEMENT

MISM 5203 DATA MINING

DATE: AUGUST, 2014   **TIME: 2 HOURS**

INSTRUCTIONS: Answer Question One and Any Other Two Questions

QUESTION ONE: [20 Marks]

a)   Describe the meaning of the following terms in the context of data mining and warehousing

      i.   Data mining   (2 Marks)

      ii.   Information gain   (2 Marks)

b)   Briefly explain any five data mining tasks in information systems and management (5 Marks)

c)   Briefly explain the following clustering algorithms as used in data mining. Give one example for each case   (4 Marks)

      (i)   Hierarchical approach

      (ii)   Model-based approach

d)   Explain the difference between supervised learning and unsupervised learning. Give one example for each case   (4 Marks)

e)   The following table shows different types of animals

Name	Blood Type	Give Birth	Can Fly	Live in Water	Class
human	warm	yes	no	no	mammals
python	cold	no	no	no	reptiles
salmon	cold	no	no	yes	fishes
whale	warm	yes	no	yes	mammals
frog	cold	no	no	sometimes	amphibians
komodo	cold	no	no	no	reptiles
bat	warm	yes	yes	no	mammals
pigeon	warm	no	yes	no	birds

f)   Use the above table to determine the coverage and accuracy of the following rule   (2 Marks)

(blood type=warm) → mammals

g)   Briefly explain the meaning the term 'model' as used in data mining   (1 Marks)

QUESTION TWO (15 MARKS)

a)   State and explain any two motivations of data mining   (2 Marks)

b)   Consider the following table.

		Features	
	Height	Weight	Age
Tom	5ft	60	35
peter	6 ft	75	49

Compute city block distance between the two people   (2 Marks)

c)   Distinguish between pre-prunning and post pruning and their importance in decision tree learning   (3 Marks)

d)   Briefly explain methods of choosing an attribute for partitioning the tree during decision tree learning   (4 Marks)

e)   Explain briefly 2 conditions for stopping building of tree during decision tree learning

(2 Marks)

f)   Describe two types of data warehouses   (2 Marks)

QUESTION THREE (15 MARKS)

a)   Describe the meaning of the following data mining techniques and explain how they can be applied in business enterprises

i) Association mining   (4 Marks)

ii. Sequence mining   (4 Marks)

b) Briefly describe any three major data mining challenges in the context of information systems and management.   (3 Marks)

c) The following table shows a list of transaction items.

TID	Items
1	Bread, Milk
2	Bread, Diaper, Beer, Eggs
3	Milk, Diaper, Beer, Coke
4	Bread, Milk, Diaper, Beer
5	Bread, Milk, Diaper, Coke

i) State and explain two metrics of evaluating association patterns that can be generated from the above data   (2 Marks)

ii) Use the identified metrics to evaluate the following association pattern (2 Marks)

$$\{Milk, Diaper\} \Rightarrow Beer$$

QUESTION FOUR (15 MARKS)

a) Explain the meaning of the following data mining terms

 i) Data Warehousing   (2 Marks)

 ii) Data mart   (2 Marks)

 iii) Operational data store   (2 Marks)

b) Briefly explain five properties of a data warehouse   (5 Marks)

c) State and explain four Metrics of evaluating interestingness of patterns during post processing of data mining task   (4 Marks)