

**B.Tech (Computer Engineering), 7<sup>th</sup> Semester Annual Examination, 2018**

**DATA MINING**

**Paper No: CEN – 701**

**Time: 3 Hours**

**Max. Marks: 60**

**Instruction: Attempt any 2 parts from each question. Each part of the question carries 6 marks.**

- 1.(a) CO – 1 What are the different issues in Data Mining?
- (b) CO – 1 What are the major challenges of mining a huge amount of data (such as millions of tuples) in comparison with mining a small amount of data (such as a few hundred tuple data set)?
- (c) CO – 1 What are the factors that have to be considered when data mining is to be performed on the following data repository?
- i) Text Data
  - ii) Web Data
- 2.(a) CO – 2 What are the various dimensionality reduction techniques used in data mining. Explain any one of the techniques.
- (b) CO – 2 Suppose that the data for analysis includes the attribute age. The age values for the data tuples are (in increasing order) 13, 15, 16, 16, 19, 20, 20, 21, 22, 22, 25, 25, 25, 25, 30, 33, 33, 35, 35, 35, 35, 36, 40, 45, 46, 52, 70.
- (a) What is the mean of the data? What is the median?
  - (b) What is the mode of the data? Comment on the data's modality (i.e., bimodal, trimodal, etc.).
  - (c) What is the midrange of the data?
  - (d) Can you find (roughly) the first quartile (Q1) and the third quartile (Q3) of the data?
  - (e) Give the five-number summary of the data.
  - (f) Show a boxplot of the data.
- (c) CO – 2 In many problems identifying an outlier is an important activity? Suggest means by which you can perform outlier detection. Also suggest different applications where we can apply such activity.

3. (a)	CO – 3	<u>Transaction Id</u>	<u>Items bought</u>
		1	Milk, Beer, Diaper
		2	Bread, Butter, Milk
		3	Milk, Diaper, Cookies
		4	Bread, Butter, Cookies
		5	Beer, Cookies, Diaper
		6	Milk, Diaper, Bread, Butter
		7	Bread, Butter, Diapers
		8	Beer, Diapers
		9	Milk, Diapers, Bread, Butter
		10	Beer, Cookies

Consider the market basket transaction shown above

- a) What is the maximum number of association rules that can be extracted from this data.(including rules that have zero support)
- b) What is the maximum size of the frequent items that can be extracted?

(b)

- c) Find an item set (of size 2 or larger) that has the largest support.  
 d) Find a pair of items, a and b such that the rules  $\{a\} \rightarrow \{b\}$  and  $\{b\} \rightarrow \{a\}$  have the same confidence.

(b) CO-3 Given a simple-transactional database, find FP-tree for this database is support threshold is 3

TID	Items
1	a,b,c,d
2	a,c,d,f
3	c,d,e,g,a
4	a,d,f,b
5	b,c,g
6	d,f,g
7	a,b,g
8	c,d,f,g

- (c) CO-3
- i) Give a short example to show that items in a strong association rule may actually be *negatively correlated*.
- ii) What is closed and max frequent mining? Using an example show that how frequent and closed items are mined.

4. a) CO-4 What is the difference between Information Gain, Gain Ratio and Gini Index? Using decision tree classifier find the root of the decision tree for the following training data. Play is the output .

Outlook	Temperature	Humidity	Windy	Play?
sunny	hot	high	false	No
sunny	hot	high	true	No
overcast	hot	high	false	Yes
rain	mild	high	false	Yes
rain	cool	normal	false	Yes
rain	cool	normal	true	No
overcast	cool	normal	true	Yes
sunny	mild	high	false	No
sunny	cool	normal	false	Yes
rain	mild	normal	false	Yes
sunny	mild	normal	true	Yes
overcast	mild	high	true	Yes
overcast	hot	normal	false	Yes
rain	mild	high	true	No

- (b) CO – 4 Explain the working principle of Support vector machine. Why SVM is sometimes considered superior to other classification algorithm on some problems.
- (c) CO – 4 Given the following observations, classify a Red Domestic SUV in to one of the class (Stolen) using Naïve Bayes Classification.

Example No.	Color	Type	Origin	Stolen?
1	Red	Sports	Domestic	Yes
2	Red	Sports	Domestic	No
3	Red	Sports	Domestic	Yes
4	Yellow	Sports	Domestic	No
5	Yellow	Sports	Imported	Yes
6	Yellow	SUV	Imported	No
7	Yellow	SUV	Imported	Yes
8	Yellow	SUV	Domestic	No
9	Red	SUV	Imported	No
10	Red	Sports	Imported	Yes

- 5.(a) CO – 5 Describe how to compute the dissimilarity between objects described by the following types of variables.
- (i) Numerical (interval-scaled) variables
  - (ii) Asymmetric binary variables
  - (iii) Categorical variables
  - (iv) Ratio-scaled variables
  - (v) Nonmetric vector objects
- (b) CO – 5 Describe each of the following clustering algorithms in terms of the following criteria: (i) shapes of clusters that can be determined; (ii) input parameters that must be specified; and (iii) limitations.
- (a) k-medoids
  - (b) AGNES
  - (c) BIRCH
  - (d) DBSCAN

What is the basic difference between Density based clustering methods and Partition based clustering methods?

(c) CO - 5

Consider a data set of twelve objects as follows:

Data Point	X value	Y value
X1	3	7
X2	4	8
X3	3	9
X4	4	9
X5	6	10
X7	5	12
X8	7	11
X9	8	12
X10	11	16
X11	10	15
X12	9	10

Considering  $K=2$ , suppose X3 and X11 are taken as two initial medoids.

- Calculate the cost of selecting these two points as medoids.
- What is the cost of shifting the medoids to X4 and X10 from its current selection?
- What will be the data points after the first iteration in both the clusters after shifting the medoids has taken place.



CEN-704

B.Tech.(Computer Engineering), VII Semester Examination 2018

Parallel and Distributed Computing

Paper No: CEN-704

Time:-03 Hours

Max. Marks:-60

Write your Roll No. on top immediately on receipt of question paper.

Attempt all questions by attempting any two parts from each question. All questions carry equal marks.

CO1**Q1(a)**

Three enhancements with the following speed ups are proposed for a new architecture: Speedup<sub>1</sub>=30, speedup<sub>2</sub>= 20, and speedup<sub>3</sub> =15. Only one enhancement is usable at a time. Assume the enhancements can be used 25%, 35% and 10% of the time for enhancements 1, 2 and 3 respectively. For what fraction of the reduced execution time is no enhancement in use? If only two enhancements are to be used then which two enhancements should be used? 06

**Q1(b)**

(i). An application program is executed on a nine computer cluster. A benchmark program takes time T on this cluster. Further 25% of T is time in which the application is run simultaneously on all nine computers. The remaining time, the application has to run on a single computer. Calculate the effective speedup under the aforementioned condition as compared to executing the program on a single computer. Also calculate the percentage of code that has been parallelized in the preceding program. 03

(ii) Let a be the percentage of program code that can be executed simultaneously by n computers in a cluster, each computer using a different set of parameters or initial conditions. Assume that the remaining code must be executed simultaneously by a single processor. Each processor has an execution rate of x MIPS. Determine an expression for the effective MIPS rate when using the system for exclusive execution of this program in terms of a, n and x. 03

**Q1**

(c). Consider a computer which can execute a program in two operational modes; regular mode versus enhanced mode with a probability distribution of {A, 1-A} respectively. If A varies between a and b and  $0 \leq a \leq 1$ , derive an expression for the average speedup factor using harmonic mean concept. 06

CO2

**Q2(a)** Describe Branch prediction techniques with help of suitable examples and diagrams. 06

**Q2(b)(i)** Consider the following

N= no. of instructions to be executed

M= No. of segments in pipeline

P= probability that a given instruction is a unconditional branch instruction

Q= probability that a given instruction is a conditional branch instruction

R= probability that a given conditional branch instruction will cause branching.

Calculate the followings

- Speedup
- Throughput
- efficiency
- Average no. of Instructions executed per Instruction cycle.

04

**Q2(b)**

(ii) Derive an expression for optimal number of stages in a pipeline.

02

**Q2(c)** Three functional pipelines f1, f2 and f3 are characterized by the following reservation tables.

06

f1:

	1	2	3	4
S1	X			
S2		X		
S3			X	X

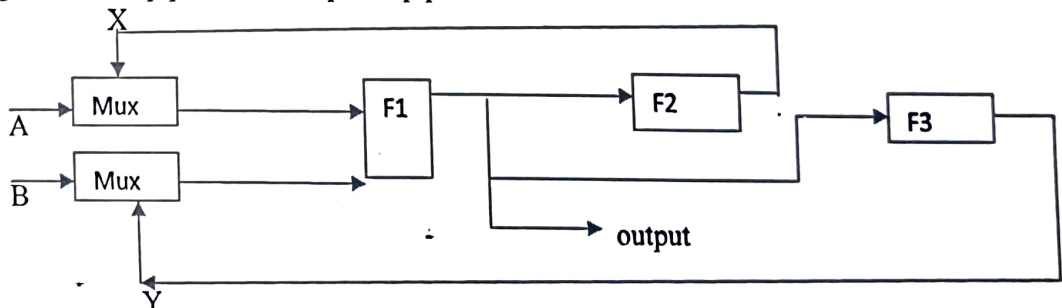
f2:

	1	2	3	4
T1	X			X
T2		X		
T3			X	

f3:

	1	2	3	4
U1	X		X	
U2				X
U3		X		

Using these three pipelines a composite pipeline is formed as below.



Each task going through this composite pipeline in the following order: f1 first, f2 and f3 next, f1 again and then the output is obtained. The dual multiplexer selects a pair of inputs (A,B) or (X,Y) and feeds them into the input of f1.

- Draw the state transition diagram for collision free scheduling
- List all simple cycle and greedy cycle
- Find MAL reservation table.

### CO3

Q3 (a) Design and verify SIMD algorithm for NXN Matrix multiplication with computational complexity as  $O(N^2)$ . 06

Q3 (b) What is PRAM model? Design and verify PRAM Algorithm for EREW NXN matrix multiplication with computational complexity of  $O(N)$ . 06

Q3(c) Parallelize Quick Sort Algorithm. Compute its speed up with respect to its sequential algorithm. 06

### CO4

Q4(a). What is MPI program? Write an MPI program for computing the values of  $N!$  06

Q4(b) What do you mean by Open Mp program ? Write an open MP program to compute the value of PI 06

Q4(c) Explain the following terms (i) GPU (ii) CUDA. Explain how a CUDA program is written with help of suitable example. 06

### CO5

Q5. Describe in details any two of the following with the help of suitable examples and diagrams.

(a) Mobile Agent

(b) Object request Broker

(c) Matrix Logical clock

2x6=12

**B.Tech. (Computer Engineering)**  
**VII Semester Examinations, 2018**  
**Artificial Intelligence**  
**Paper No. CEN-703**

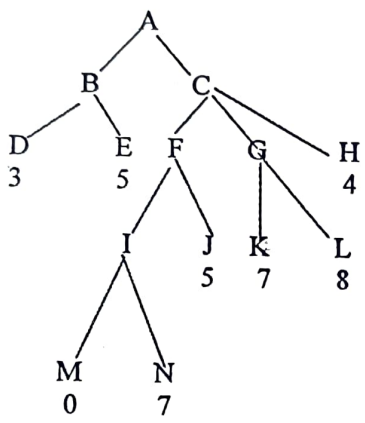
**Maximum Time: 3 hours.**

**Maximum Marks: 60**

**Note: Attempt all questions. All questions carry equal marks. Assume suitable missing data, if any.**

			Marks
1	<b>Q No. 1a</b>	Consider the following paragraph:	6
	<b>CO1</b>	<p>Sam, Clyde and Oscar are elephants. The following facts about them are known:</p> <ol style="list-style-type: none"> <li>1. Sam is pink.</li> <li>2. Clyde is gray and likes Oscar.</li> <li>3. Oscar is either pink or gray (but not both) and likes Sam.</li> </ol> <p>Use resolution refutation to prove that a gray elephant likes a pink elephant; that is, prove <math>(\exists x, y)[\text{Gray}(x) \wedge \text{Pink}(y) \wedge \text{likes}(x, y)]</math>.</p>	
	<b>Q No. 1b</b>	Use resolution refutation on a set of clauses to prove that there is a - green object if following are given	6
	<b>CO1</b>	<ul style="list-style-type: none"> <li>• If pushable objects are blue, then nonpushable ones are green.</li> <li>• All objects are either blue or green but not both.</li> <li>• If there is a nonpushable object, then all pushable ones are blue.</li> <li>• Object 01 is pushable.</li> <li>• Object 02 is not pushable.</li> </ul> <ol style="list-style-type: none"> <li>1. Convert these statements to expressions in first-order predicate calculus.</li> <li>2. Convert the preceding predicate- calculus expressions to clause form.</li> <li>3. Combine the preceding clause form expressions with the clause form of the negation of the statement to be proved, and then show the steps used in obtaining a resolution refutation.</li> </ol> <p style="text-align: center;">OR</p>	
	<b>Q No.1b'</b>		
	<b>CO1</b>	Explain Means Ends Analysis search technique with a suitable example.	6
2	<b>Q No. 2a</b>	Differentiate between Informed and Uninformed Search Techniques. When would Best- first search be worse than simple breadth- first search? Explain with a suitable example.	6
	<b>CO2</b>		
	<b>Q No. 2b</b>	Using a suitable example, illustrate steps of A* Search. Why is A* search better than Best First Search.	6
	<b>CO2</b>	OR	



	Q No.2b' CO2	Define Production system. List some problem characteristics and also explain the Requirements of a good control strategy.	6
3	Q No. 3a CO3	Differentiate between forward and backward reasoning. What factors determine the choice of direction for a particular problem?	6
	Q No. 3b CO3	Write the steps of alpha- beta pruning algorithm with an example. OR	6
	Q No.3b' CO3	Perform minimax on the following tree: 	6
4	Q No. 4a CO4	What is an Expert System? How is it different from Conventional Computer System. List some major Applications of Expert System. Draw a Block diagram showing main component of a typical Expert System.	6
	Q No. 4b CO4	List and explain all the phases of Natural Language Processing with some suitable examples.	6
5	Q No. 5 CO5	Write short note on the following (Any three.) (i) Reasoning using frames. (ii)Winston's Learning Program. (iii) Application of Machine Learning. (iv)Parallel relaxation process in Hopfield network. (v) Case-Based Reasoning.	[4*3=12]

CEN-802

Roll No.....

B.Tech. (Computer Engineering) 8<sup>th</sup> Semester Examination, 2016

Artificial Intelligence

Paper No: CEN-802

Time: 3 Hours

Max Marks: 60

*(Write your Roll No. on the top immediately on the receipt of this Question paper.)*

*Note: Attempt all Questions, Assume suitable missing data, If any.*

1. (a) We are given the following paragraph:

[6]

Tony, mike and john belong to Alpine Club. Every member of the Alpine Club is either a skier or a mountain climber or both. No mountain climber likes rain, and all skiers like snow. Mike dislikes whatever Tony likes and Likes whatever Tony dislikes. Tony likes rain and snow.

Represent this information by Predicate- calculus sentences in such a way that you can represent the question "Who is a member of Alpine Club who is a mountain climber but not a skier?" as a predicate- calculus expression. Use resolution refutation with answer extraction to answer it.

Or

(a') Use resolution refutation on a set of clauses to prove that there is a green object if we are given:

[6]

- If pushable objects are blue, then nonpushable ones are green.
- All objects are either blue or green but not both.
- If there is a nonpushable object, then all pushable ones are blue.
- Object 01 is pushable.
- Object 02 is not pushable.

1. Convert these statements to expressions in first order predicate calculus.
2. Convert the preceding predicate- calculus expressions to clause form.
3. Combine the preceding clause form expressions with the clause form of the negation of the statement to be proved, and then show the steps used in obtaining a resolution refutation.

(b) What are the disadvantages of Steepest Hill climbing Search Procedure? Using a suitable search tree, illustrate that these drawbacks are eliminated in Best First Search. [6]

2. (a) Define Production system. Discuss some problem characteristics and also explain the requirements of a good control strategy. [6]

(b) Using a suitable example, illustrate steps of A\* Search. Why is A\* search better than Best First Search. [6]

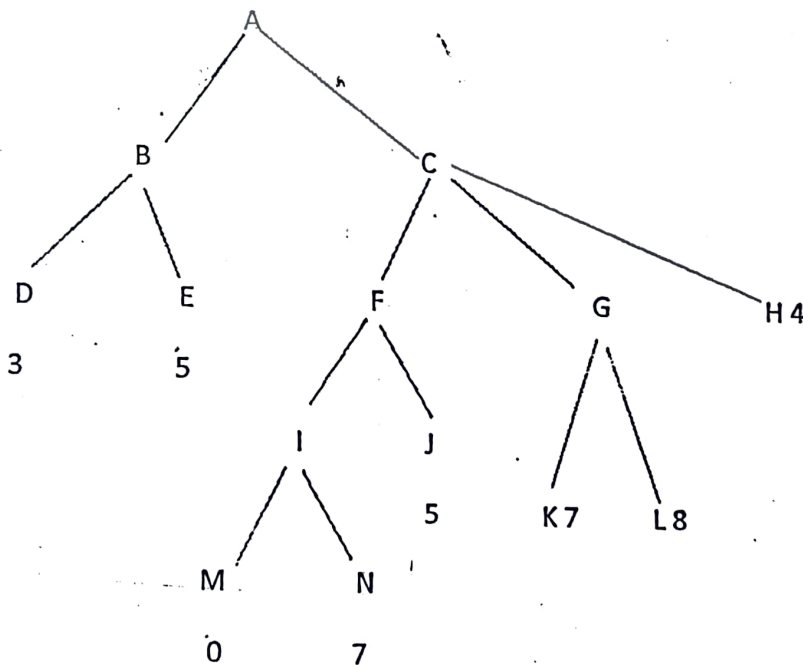
Or

(b') Differentiate between forward and backward reasoning. What factors determine the choice of direction for a particular problem? [6]

3. (a) Show a Conceptual Dependency(CD) representation of the sentence: [6]

- (i) John ate salad with a fork.
- (ii) While going home Sachin saw Brett Lee.
- (iii) Farmer fertilized the field.

(b) Perform minimax on the following tree- [6]



Or

(b') Write the steps of alpha- beta pruning algorithm with an example. [6]

4. (a) Define Machine Learning. Discuss its various applications and give classification of machine learning tasks. [6]

(b) List and explain all the phases of Natural Language Processing with some suitable examples. [6]

Or

(b') Define Expert system. How Expert system is different from conventional Computer System and also list some applications of expert system. [6]

5. Write short note on the following (Any three.)

[4x3=12]

(i) Reasoning using Frames.

(ii) Reasoning using Scripts.

(iii) Non Monotonic Reasoning.

(iv) Types of Learning.

(v) Training Problem in Neural Network.