Department of B. Sc. (Data Science)

APCCE / Class: SYDS / SEM-III / SUB: Data Warehouse and Mining/ DUR 2 ½ Hrs. / Mks: 75 / DT. _____

Note:1)All questions are compulsory.
2) Figures to the right indicate marks.
3) Writing of sub-questions is not allowed and try to accommodate whole QP in single page.

Q1.	Attempt any THREE:		[15]
	a.	Distinguish between operational and informational systems.	
	b.	Name at least six characteristics or features of data warehouse.	
	c.	What is mean by data fusion?	
	d.	Define Data Warehouse.	
	e.	List and describe three major reasons why metadata is vital for end-users.	
	f.	Name four distinguishing characteristics of the data warehouse architecture. Describe each briefly.	
Q2.	Attempt any THREE:		[15]
	a.	What is the STAR schema? What are the component tables?	
	b.	State any three advantages of the STAR schema.	
	c.	What are aggregate fact tables? Why are they needed? Give an example.	
	d.	Define initial load, incremental load, and full refresh.	
	e.	Name five types of the major transformation tasks. Give an example for each.	
	f.	Describe with examples snapshot and transaction fact tables. How are they related?	
Q3.	Attempt any THREE:		[15]
	a.	List of any five different data mining applications.	
	b.	Write a short note on ARFF.	
	c.	Explain the different phases of data mining process in brief.	
	d.	What do you mean by normalization and standardization in data transformation?	
	e.	What is the decision tree classifier? Give an example.	
	f.	Difference between Data Mining and Machine learning.	
Q4.	Attempt any THREE:		[15]
	a.	Write short notes on Euclidean distance.	
	b.	Explain the agglomerative hierarchical clustering algorithm in brief.	
	c.	What do you mean by web usage mining?	
	d.	List the applications of cluster analysis.	
	e.	Write a short note on web crawler.	
	f.	Explain resemblance and containment in web content mining.	
Q5.	Attempt any THREE:		[15]
	a.	Define Association Rule Mining.	
	b.	Explain working of the Apriori algorithm.	
	c.	Explain the approaches for transaction database storage in Association Mining.	
	d.	What do you mean by the closed and maximal itemsets in the Apriori algorithm?	
	e.	Explain the concept of mining frequent patterns without candidate generation (FP Growth).	
	f.	Give the representations of Items for Association Mining.	

—X—X—