VCD/ 28032023 **S.Y.B.Sc.DS** **SEM-IV Big Data** **2½ hrs** **75 Marks**

- All questions are **compulsory.**
- All questions **carry equal Marks.**
- Draw **neat, labelled diagrams** whenever necessary.

Q.1	**Attempt all the questions.**		15
	a.	Explain Outlining steps in the life cycle.	
	b.	Explain the Examples of data visualization	
	c.	Differentiate Traditional analytics with Big Data analytics.	
	d.	How to exploit data to optimize Data-Driven Decisions?	
	e.	Explain KDD.	
	f.	Write a short note on structured data	
Q.2	**Write the answers of following questions (*Any three*)**		15
	a.	Explain the concept of Predictive analytics.	
	b.	Explain the concept of Supervised learning	
	c.	Explain the concept of Modelling.	
	d.	How to discover unknown groups through clustering?	
	e.	Write a short note on decision tree classification.	
	f.	Write a short note on Time Series Analysis.	
Q.3	**Write the answers of following questions (*Any three*)**		15
	a.	Explore the Hadoop ecosystem.	
	b.	Explain the Steps of MapReduce Job Execution flow.	
	c.	Explain Working of the MapReduce Architecture.	
	d.	What is RHadoop? Explain the purpose behind R and Hadoop Integration.	
	e.	Write a short note on Recommendation Framework.	
	f.	What is Streamlining? How Does Data Analytics Work?	
Q.4	**Write the answers of following questions (*Any three*)**		15
	a.	What is Data Mining? Explain the steps.	
	b.	What are Problems faced in storing unstructured data?	
	c.	What are ways to Analyse Unstructured Data?	
	d.	Write a short note on "Analysing the structure of text mining"	
	e.	Explain the business case for unstructured data mining?	
	f.	Explain the pros and cons of Text Mining.	
Q.5	**Write the answers of following questions (*Any three*)**		15
	a.	Explain how Planning and Implementing a complete Data analytics projects.	
	b.	How to transform business objectives to analytics projects? Explain.	

	c.	How to identify the performance metrics targets? Explain.	
	d.	Write a short note on finding core data sets.	
	e.	Write a short on preparation of data for analysis	
	f.	Write a short on executing the model	