Reg.No.:												

VIVEKANANDHA COLLEGE OF ENGINEERING FOR WOMEN
[AUTONOMOUS INSTITUTION AFFILIATED TO ANNA UNIVERSITY, CHENNAI]
Elayampalayam – 637 205, Tiruchengode, Namakkal Dt., Tamil Nadu.

Question Paper Code : 60020

B.E. / B.Tech. DEGREE END-SEMESTER EXAMINATIONS – NOV. / DEC.2024
Fifth Semester
Information Technology
U19ITV31 - DATA SCIENCE
(Regulation 2019)

Time: Three Hours Maximum: 100 Marks

Answer ALL the questions

Knowledge Levels (KL)	K1 – Remembering	K3 – Applying	K5 - Evaluating
	K2 – Understanding	K4 – Analyzing	K6 - Creating

PART – A

(10 x 2 = 20 Marks)

Q.No.	Questions	Marks	KL	CO
1.	How does Hadoop handles large datasets?	2	K1	CO1
2.	Outline the purpose of Data preparation phase. How it is done?	2	K2	CO1
3.	Paraphrase the importance of data sampling while dealing with massive data.	2	K2	CO2
4.	Define the term partition tolerance.	2	K1	CO2
5.	Give an example for distributed data storage.	2	K2	CO3
6.	Paraphrase about OWASP.	2	K2	CO3
7.	State any two real- time scenarios that can be stored as graphs in Graph database.	2	K3	CO4
8.	How can a system achieve eventual consistency and how does it relates to the CAP Theorem?	2	K2	CO4
9.	What is Tokenization?	2	K1	CO5
10.	Differentiate between text analytics and text mining.	2	K2	CO5

PART – B

(5 x 13 = 65 Marks)

Q.No.		Questions	Marks	KL	CO
11.	a)	With a neat block diagram, explain the Data science process life cycle.	13	K1	CO1
		(OR)			
	b)	Describe a data analysis use case using Hadoop framework. What are all the challenges associated with it and explain how to mitigate them.	13	K2	CO1
12.	a)	What are the various machine learning techniques that can be used in handling big data? Explain.	13	K2	CO2
		(OR)			
	b)	Discuss any two effective storage techniques for sorting large data. Analyze the trade-off between accessing speed vs storage.	13	K3	CO2
13.	a)	With a neat block diagram, explain the core components of Hadoop architecture.	13	K3	CO3
		(OR)			
	b)	Outline the working mechanism of SPARK framework. Discuss in brief about its characteristics and functions.	13	K3	CO3
14.	a)	Explain the significance of ACID property. With two suitable examples of database systems that emphasize ACID properties and one that aligns more with the CAP theorem.	13	K3	CO4
		(OR)			
	b)	Discuss in detail about the features, advantages and disadvantages of NoSQL.	13	K4	CO4
15.	a)	Enumerate in detail about the various text mining techniques.	13	K4	CO5
		(OR)			
	b)	i. What is lemmatization? Explain with an example.	7	K2	CO5
		ii. Write notes on Decision tree classifier.	6	K3	CO5

PART – C

(1 x 15 = 15 Marks)

Q.No.		Questions	Marks	KL	CO
16.	a)	What is an Exploratory data analysis? Explain in detail about the steps involved in performing Exploratory data analysis.	15	K3	CO1
		(OR)			
	b)	i. With an example, explain how graph databases will be of use in healthcare/ social media applications.	8	K3	CO4
		ii. Outline a methodology to classify Reddit posts.	7	K3	CO5