Reg. No. | | | | | | | | | | | | | |

B.E / B.TECH. DEGREE EXAMINATION, MAY 2023

Eighth/Sixth Semester

CS18008 BUSINESS INTELLIGENCE

(Common to CS & AD)

(Regulation 2018)

TIME: 3 HOURS **MAX. MARKS: 100**

CO 1 Students will be able to apply the ETL concepts, tools and techniques to perform Extraction, Transformation and Loading of data

CO 2 Students will be able to summarize the usable data by using various reporting concepts, techniques/tools, and use charts, tables for reporting in BI

CO 3 Students will be able to use Analytics concepts like data mining, Exploratory and statistical techniques for predictive analysis in Business Intelligence

CO 4 Students will be able to demonstrate application of concepts in BI.

CO 5 Students will be able to analyze and carry out an empirical evaluation of different algorithms on a problem formalization, and state the conclusions that the evaluation supports

PART- A (10 x 2 = 20 Marks)

(Answer all Questions)

		CO	RBT LEVEL
1.	Why is the web considered so important for decision support?	**1**	**2**
2.	What are the major similarities and differences of DSS and BI?	**1**	**2**
3.	When do you go for Pie chart visualization?	**2**	**4**
4.	What do you mean by colour pollution?	**2**	**4**
5.	Why do you need k-fold cross validation methodology?	**3**	**4**
6.	Mention the taxonomy of ANN Learning algorithms and architectures.	**3**	**3**
7.	What are hubs and authorities? What is the HITS algorithm?	**4**	**2**
8.	Identify the pros and cons of choosing a free text mining tool over a commercial tool.	**4**	**4**
9.	When do you go for Decomposition trees visualization method?	**5**	**4**
10.	What do you mean by Navigable reports?	**5**	**2**

PART- B (5 x 14 = 70 Marks)

		Marks	CO	RBT LEVEL
11. (a)	Norfolk Southern Corporation is one of the premier transportation companies in the United States. With a reputation as the "Thoroughbred of Transportation," its Norfolk Southern Railway subsidiary operates approximately 20,000 route miles, which crisscross 22 states and the District of Columbia. But Norfolk Southern is not content to reflect on its rich 182-year rail industry history. Working to achieve its vision of being "the safest, most customer-focused and successful transportation company in the world," the company is galloping ahead of the competition with new business models and supporting technology. A. How are information systems used at Norfolk Southern to support decision making? B. What type of information is accessible through the visualization applications? C. What type of information support is provided through accessNS? D. Can the same data warehouse be used for business intelligence and optimization applications?	**(14)**	**1**	**2**
	(OR)			
(b)	Examine how does data warehouse, business analytics and business performance management play a major role in the architecture of Business Intelligence	**(14)**	**1**	**2**
12. (a)	The Iris Dataset contains four features (length and width of sepals and petals) of 50 samples of three species of Iris (Iris setosa, Iris virginica and Iris versicolor). These measures are used to classify the species. The dataset is often used in data mining, classification and clustering examples and to test algorithms. Information about the original paper and usages of this dataset can be found in the UCI Machine Learning Repository.			

	A	B	C	D	E
1	sepal_length	sepal_width	petal_length	petal_width	species
2	5.1	3.5	1.4	0.2	setosa
3	4.9	3	1.4	0.2	setosa
4	4.7	3.2	1.3	0.2	setosa
5	4.6	3.1	1.5	0.2	setosa
6	5	3.6	1.4	0.2	setosa
7	5.4	3.9	1.7	0.4	setosa
8	4.6	3.4	1.4	0.3	setosa
9	5	3.4	1.5	0.2	setosa

Visualize the relationship for the information given below using Python

A) Plot a graph with length of petals as the x-axis and the breadth of petals as the y-axis **(7) 2 4**

B) Plot a scatter plot graph with both sepals and petals with length as the x-axis and breadth as the y-axis **(7) 2 4**

(OR)

(b) Differentiate Scientific Visualization from Information Visualization. Also identify the principles involved in improving the vision **(14) 2 4**

13. (a) Apply Apriori algorithm to illustrate the sales transactions in a grocery store with suitable examples **(14) 3 3**

(OR)

(b) Demonstrate how do neural networks help the Forum of International Irregular Network Access (FIINA) to reduce Telecommunications fraud **(14) 3 3**

14. (a) Examine how textual data can be captured automatically using Web-based technologies **(14) 4 4**

(OR)

(b) In the 1880's, Francis Galton was developing ways to quantify the heritability of traits. As part of this work, he collected data on the heights of adult children and their parents. Entries were deleted for those children whose heights were not recorded numerically by Galton, who sometimes used entries such as "tall", "short", "idiotic", "deformed" and so on**.** A data frame with 898 observations on the following variables

- **family** : a factor with levels for each family
- **father** : the father's height (in inches)
- **mother** : the mother's height (in inches)
- **sex** : the child's sex: F or M
- **height** : the child's height as an adult (in inches)
- **nkids** : the number of adult children in the family, or, at least, the number whose heights Galton recorded.

Family	Father	Mother	Gender	Height	Kids
1	78.5	67.0	M	73.2	4
1	78.5	67.0	F	69.2	4
1	78.5	67.0	F	69.0	4
1	78.5	67.0	F	69.0	4
2	75.5	66.5	M	73.5	4
2	75.5	66.5	M	72.5	4

Visualize the following using R packages

(a) Use Histogram to display the amount of people that fall into various different height ranges **(7) 4 4**

(b)Use Scatterplot to show Galton data on heights of parents and adult child **(7) 4 4**

15. (a) Illustrate with a suitable framework to evaluate the changes taking place in BI technology **(14) 5 3**

(OR)

(b) Demonstrate how does BI Search offer a number of promising benefits to business intelligence? **(14) 5 3**

PART- C (1 x 10 = 10 Marks)

(Q.No.16 is compulsory)

		Marks	CO	RBT LEVEL
16.	Design a Multilayer Perceptron ANN structure for Credit Card Fraud Detection Problem which includes modeling past credit card transactions with the knowledge of the ones that turned out to be fraud. This model has to be used to identify whether a new transaction is fraudulent or not. Also it should detect 100% of the fraudulent transactions while minimizing the incorrect fraud classifications.	**(10)**	**3**	**5**

****