

Enrollment No.....



**Faculty of Engineering**  
**End Sem (Even) Examination May-2022**

**CB3EL03 Data Mining & Analytics**

Programme: B.Tech.

Branch/Specialisation: CSBS

**Duration: 3 Hrs.****Maximum Marks: 60**

Note: All questions are compulsory. Internal choices, if any, are indicated. Answers of Q.1 (MCQs) should be written in full instead of only a, b, c or d.

- Q.1 i. Data mining provides the enterprise with \_\_\_\_\_. 1  
 (a) Knowledge (b) Intelligence  
 (c) Quantity (d) Memory
- ii. Photo tagging is an example of \_\_\_\_\_. 1  
 (a) Statistics (b) Classification  
 (c) Machine learning (d) Database management system
- iii. \_\_\_\_\_ normalization is not so efficient in handling the outliers. 1  
 (a) Min-max (b) Z-score  
 (c) Decimal scaling (d) None of these
- iv. Simple random sampling methods may have very poor performance in the presence of \_\_\_\_\_. 1  
 (a) Skew (b) Symmetric  
 (c) Reduced (d) None of these
- v. \_\_\_\_\_ step of apriori algorithm generates k+1 item set from k- itemsets by joining each item with itself. 1  
 (a) Join (b) Pre prune  
 (c) Post prune (d) All of these
- vi. Class0: 5, Class1: 5 indicate the \_\_\_\_\_ degree of impurity. 1  
 (a) Low (b) Minimum  
 (c) High (d) Multi
- vii. What is the meaning of the testing of the hypothesis? 1  
 (a) It is a method of making a significant statement.  
 (b) It is a rule for acceptance or rejection of the hypothesis of the research problem.  
 (c) It is a significant estimation of the problem.  
 (d) None of these

P.T.O.

[2]

- viii. The Newton-Raphson method of finding roots of nonlinear equations falls under the category of which method? **1**  
 (a) Random (b) Open (c) Graphical (d) Bracketing
- ix. A rise in prices before the festival is an example of \_\_\_\_\_. **1**  
 (a) Irregular trend (b) Secular trend  
 (c) Cyclic trend (d) Seasonal trend
- x. Which of the following is not a technique used in smoothing time series? **1**  
 (a) Nearest Neighbour Regression  
 (b) Locally weighted scatter plot smoothing  
 (c) Tree-based models like (CART)  
 (d) Smoothing splines
- Q.2 i. Define the term data mining and write its application in the real world. **2**  
 ii. What are the challenges of data mining? **3**  
 iii. What are the stages of data mining? **5**
- OR iv. Write a short note on OLAP. **5**
- Q.3 i. Explain the techniques used for data transformation. **4**  
 ii. Using the data for marks: **6**  
 13, 15, 16, 19, 20, 21, 22, 22, 25, 25, 25, 25, 30, 33, 33, 35, 35, 35, 35, 36, 40, 45, 46, 52, 70  
 Apply smoothing by bin mean, median and boundaries method.  
 Take a bin depth of 3.
- OR iii. What defines a data mining task? **6**
- Q.4 i. Explain the antimonotone property of the apriori algorithm. **2**  
 ii. What are the advantages of a decision tree? Explain design issues of decision tree induction. **3**  
 iii. Define the zero-probability problem. What is the solution to this problem? Explain by an example. **5**
- OR iv. Explain the attribute selection measures. **5**
- Q.5 Attempt any two:  
 i. Explain the Wald test and LR test. Write any two difference points between Wald and LR test. **5**

[3]

- ii. Explain the generalized linear model. **5**  
 iii. What is trend analysis? Explain its four movements with an example. **5**
- Q.6 i. Explain the working of the Holt-Winter smoothing method used for time series data. **4**  
 ii. Answer the following questions for ARMA: **6**  
 (a) Define ARMA.  
 (b) How to choose an ARMA model?  
 (c) Why ARMA model be considered particularly useful for financial time series data?
- OR iii. Answer the following questions: **6**  
 (a) Differentiate between prescriptive analytics and descriptive analytics.  
 (b) Name the tests used for trend and seasonality.  
 (c) Define Yule-Walker estimation for AR processes?

\*\*\*\*\*

## Marking Scheme

### CB3EL03 Data Mining & Analytics

Q.1	i.	Data mining provides the enterprise with _____.		1
		(b) Intelligence		
	ii.	Photo tagging is an example of _____.		1
		(c) Machine learning		
	iii.	_____ normalization is not so efficient in handling the outliers.		1
		(a) Min-max		
	iv.	Simple random sampling methods may have very poor performance in the presence of _____.		1
		(a) Skew		
	v.	_____ step of apriori algorithm generates k+1 item set from k-itemsets by joining each item with itself.		1
		(a) Join		
Q.2	vi.	Class0: 5, Class1: 5 indicate the _____ degree of impurity.		1
		(c) High		
	vii.	What is the meaning of the testing of the hypothesis?		1
		(b) It is a rule for acceptance or rejection of the hypothesis of the research problem.		
	viii.	The Newton-Raphson method of finding roots of nonlinear equations falls under the category of which method?		1
		(b) Open		
	ix.	A rise in prices before the festival is an example of _____.		1
		(d) Seasonal trend		
	x.	Which of the following is not a technique used in smoothing time series?		1
		(c) Tree-based models like (CART)		
Q.2	i.	Definition of Data mining	1 mark	2
		Applications	1 marks	
	ii.	Any six challenges	(0.5 mark * 6)	3
	iii.	Five stages	(1 mark * 5)	5
OR	iv.	Definition of OLAP	1 mark	5
		Applications	(0.5 mark * 2)	
		Features (any two)	(0.5 mark * 2)	
		Types (name only)	2 marks	
Q.3	i.	Any four transformation techniques (1 mark * 4)	4 marks	4

OR	ii.	As per solution	6 marks	6
	iii.	Data mining task		6
		As per explanation	6 marks	
Q.4	i.	Antimonotone property	2 marks	2
	ii.	Advantages of a decision tree	1 mark	3
		Design issues	2 marks	
	iii.	Definition	1 mark	5
		Solution	2 marks	
		Example	2 marks	
OR	iv.	Attribute selection measures		5
		As per explanation	5 marks	
Q.5		Attempt any two:		
	i.	Wald test	2 marks	5
		LR test	2 marks	
		Any two difference	1 mark	
	ii.	Definition	1 mark	5
		Components	2 marks	
		Limitations	1 mark	
		Advantages of GML over traditional regression	1 mark	
	iii.	Definition of trend analysis	1 mark	5
		Any four Components (1 mark * 4)	4 marks	
Q.6	i.	Working of the Holt-Winter smoothing method	4 marks	4
	ii.	(a) Define ARMA	2 marks	6
		(b) Choose an ARMA model	2 marks	
		(c) Why ARMA model considered	2 marks	
	OR	iii. Answer the following questions:		6
		(a) Differences	2 marks	
		(b) Name the tests	2 marks	
		(c) Yule-Walker estimation	2 marks	

\*\*\*\*\*