

# JK LAKSHMIPAT UNIVERSITY, JAIPUR

## INSTITUTE OF ENGINEERING AND TECHNOLOGY

End Term Examination November 2018

B. Tech in CSE (IBM), Semester V

Roll No.....

CSEBD501 : Enterprise Reporting using Business Intelligence

Time: 3 hours

Max. Marks: 80

---

### Instructions to students:

1. Do not write anything other than your roll number on question paper.
  2. Assume suitable data wherever required and mention it clearly.
  3. Writing appropriate units, nomenclature, and drawing neat sketches/schematics/ flow chart wherever, required are an integral part of the answer.
- 

Q.1A Objective type questions (10 questions @ 1 mark)

10

- i. Bayesian classifiers is
  - A. A class of learning algorithm that tries to find an optimum classification of a set of examples using the probabilistic theory.
  - B. Any mechanism employed by a learning system to constrain the search space of a hypothesis
  - C. An approach to the design of learning algorithms that is inspired by the fact that when people encounter new situations, they often explain them by reference to familiar experiences, adapting the explanations to fit the new situation.
  - D. None of these
- i. Data transformation includes which of the following?
  - A. A process to change data from a detailed level to a summary level
  - B. A process to change data from a summary level to a detailed level
  - C. Joining data from one source into various sources of data
  - D. Separating data from one source into various sources of data
- ii. The extract process is which of the following?
  - A. Capturing all of the data contained in various operational systems
  - B. Capturing a subset of the data contained in various operational systems
  - C. Capturing all of the data contained in various decision support systems
  - D. Capturing a subset of the data contained in various decision support systems
- iii. Fact tables are completely normalized or not. Justify your answer.
- iv. .... is a subject-oriented, integrated, time-variant, nonvolatile collection or data in support of management decisions.
- v. Weka data source supports ..... file format
- vi. .... is a good alternative to the star schema.
  - A) Star schema
  - B) Snowflake schema
  - C) Fact constellation

- vii. D) Star-snowflake schema  
The type of relationship in star schema is .....
- A) many to many  
B) one to one  
C) one to many  
D) many to one
- viii. Which of the following is not a kind of data warehouse application?  
A) Information processing  
B) Analytical processing  
C) Data mining  
D) Transaction processing
- ix. The star schema is composed of \_\_\_\_\_ fact table.  
A. one.  
B. two.  
C. three.  
D. four.
- x. Cognos software deals with ..... Types of data.

Q.2 Short answer type questions (4 questions @ 2.5 marks) 10

- 2.A What is literal meaning of business?  
2.B Which is bottom up approach clustering or classification. Explain  
2.C What is advantage of K nearest Neighbor over K?  
2.D What is Exploratory data analysis?

Q.3A For the following group of data: 324, 361, 400, 441, 362, 324, 291, 2134, 345, 400, 441, 324 normalize them with min = 0 and max = 100. 05

Q.3B For the above group of data Q.3A, partition them into two bins by each of the following methods: 05  
(1) equal-width partitioning,  
(2) equal-frequency partitioning.

Q.4A A random sample of beef hotdogs was taken and the amount of sodium (in mg) and calories were measured below. 05

Table 1: Calories and Sodium Levels in Beef Hotdogs

| Calories | Sodium |
|----------|--------|
| 186      | 495    |
| 181      | 477    |
| 176      | 425    |
| 149      | 322    |
| 184      | 482    |
| 190      | 587    |
| 158      | 370    |
| 139      | 322    |
| 175      | 479    |
| 148      | 375    |

|     |     |
|-----|-----|
| 152 | 330 |
| 111 | 300 |
| 141 | 386 |
| 153 | 401 |
| 190 | 645 |
| 157 | 440 |
| 131 | 317 |
| 149 | 319 |
| 135 | 298 |
| 132 | 253 |

A. construct a representation report for analysis of the data through histogram.

Q.4B Give the regression equation, and interpret the coefficients in terms of this problem of the data Q.4A. 05

Q.5A How linear regression could be used for filtering spam. 05

Q.5B Explain step performed in IBM Cognos for generating reporting on an e commerce data set. 05

Q.6A Consider the data are shown below. Identify the most important attribute for classification problem of Play (yes or no) Justify through entropy based method. 05

| outlook  | temperature | humidity | windy | play |
|----------|-------------|----------|-------|------|
| sunny    | hot         | high     | false | no   |
| sunny    | hot         | high     | true  | no   |
| overcast | hot         | high     | false | yes  |
| rainy    | mild        | high     | false | yes  |
| rainy    | cool        | normal   | false | yes  |
| rainy    | cool        | normal   | true  | no   |
| overcast | cool        | normal   | true  | yes  |
| sunny    | mild        | high     | false | no   |
| sunny    | cool        | normal   | false | yes  |
| rainy    | mild        | normal   | false | yes  |
| sunny    | mild        | normal   | true  | yes  |
| overcast | mild        | high     | true  | yes  |
| overcast | hot         | normal   | false | yes  |
| rainy    | mild        | high     | true  | no   |

Q.6B Whether the play will be 'yes' or 'no' based on inputs "temperature=cool humidity=normal windy=FALSE". Justify through entropy. 05

OR

Q.6B Perform cleaning of Q 4.A for sodium levels if storage allowed is only 16 bytes ( assume one number is equal to two bytes) 05

Q.7 Consider the training dataset shown below.

10

| A | B | Class Label |
|---|---|-------------|
| 0 | 1 | c1          |
| 0 | 0 | c2          |
| 1 | 1 | c1          |
| 0 | 1 | c1          |
| 1 | 0 | c1          |
| 0 | 0 | c2          |
| 1 | 1 | c2          |
| 0 | 0 | c2          |
| 1 | 0 | c1          |
| 1 | 0 | c2          |

Use the computed conditional probabilities to predict the class label for a test sample  $(A = 1, B = 0)$  and  $(A=1, B=1)$ .

Q.8 Consider that transaction database shown below. Identify frequent item set and arrange the items in a mall so that chance for profit making is increased.

10

| Transaction id | Items           |
|----------------|-----------------|
| t1             | {1, 2, 4, 5}    |
| t2             | {2, 3, 5}       |
| t3             | {1, 2, 4, 5}    |
| t4             | {1, 2, 3, 5}    |
| t5             | {1, 2, 3, 4, 5} |
| t6             | {2, 3, 4}       |

OR

Q.8 Compute the conditional probabilities  $P(A = 1 | C = c1)$ ,  $P(A = 0 | C = c1)$ ,  $P(B = 1 | C = c1)$ ,  $P(B = 0 | C = c1)$ ,  $P(A = 1 | C = c2)$ ,  $P(A = 0 | C = c2)$ ,  $P(B = 1 | C = c2)$ , and  $P(B = 0 | C = c2)$

10