DHARMSINH DESAI UNIVERSITY, NADIAD
FACULTY OF TECHNOLOGY
B.TECH. SEMESTER V [I.T]
SUBJECT: (IT-704) Data Analysis & Information Extraction

Examination	**: Second Sessional (Written Exam)**	**Seat No.**	**: _____**
Date	**: 10/09/2021**	**Day**	**: Thursday**
Duration	**: 45 minutes**	**Max. Marks**	**: 16**

INSTRUCTIONS:
1. Figures to the right indicate maximum marks for that question.
2. The symbols used carry their usual meanings.
3. Assume suitable data, if required & mention them clearly.
4. Draw neat sketches wherever necessary.

Q.2 Attempt ***Any Two*** from the following questions. **[8]**
(a) Define the limitation of Apriori Algorithm with suitable example. **[4]**
(b) Define various types of association rule mining with examples **[4]**
(c) Define various kind of Concept Hierarchy with proper examples. **[4]**

Q.3 (a) A database has four transactions.Let *min_sup*=2 & *min_conf* = 80%. **[4]**

TID	***Date***	***Items_bought***
T100	01/09/2003	{K, A, D, B}
T200	01/09/2003	{D, A, C, E, B}
T300	03/09/2003	{C, A, B, E}
T400	08/09/2003	{B, A, D}

List all frequent itemsets using Apriori and list at least one strong association rules for given type of patterns. E.g X ^ Y => Z.

(b) The following contingency table summarizes supermarket transaction data. **[4]**

	Hotdogs	_____ hotdogs
hamburgers	1000	500
_____ hamburgers	2000	1500

a) Suppose that the association rule "hot dogs=>hamburger" is mined. Given a minimum support threshold of 25% and a minimum confidence threshold of 50%, is this association rule strong?
b) Based on the given data, is the purchase of hot dogs independent of the purchase of hamburger? If not, what kind of correlation relationship exists between the two?

OR

Q.3 (a) For the database given in Q.3 (a) above let *min_sup*=2 & *min_conf* = 80%.
List all frequent itemsets using FP Growth and list at least strong association rules for given type of patterns. E.g X ^ Y => Z. **[4]**

(b) Numeric concept hierarchy generation by intuitive partitioning. Suppose that profits at different branches of AllElectronics for the year 2004 cover a wide range, from −$351,976.00 to $4,700,896.50. A user desires the automatic generation of a concept hierarchy for profit. For improved readability, we use the notation (l...r] to represent the interval (l,r]. For example, (−$1,000,000...$0] denotes the range from −$1,000,000 (exclusive) to $0 (inclusive). Suppose that the data within the 5th percentile and 95th percentile are between −$159,876 and $1,838,761. **[4]**

Apply 3-4-5 rule and show concept hierarchy for numeric data.