# MANIPAL INSTITUTE OF TECHNOLOGY
## MANIPAL

*A Constituent unit of MAHE, Manipal*

## VI SEMESTER B.TECH. (COMPUTER SCIENCE & ENGINEERING)
## END SEMESTER EXAMINATIONS, MAY 2023

## SUBJECT: DATA WAREHOUSE AND DATA MINING [CSE 4060]
### REVISED CREDIT SYSTEM
### (--/05/2022)

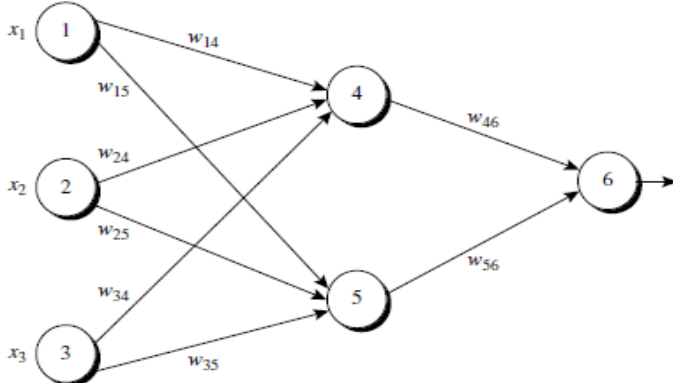Time: 3 Hours                                                                 MAX. MARKS: 50

---

**Instructions to Candidates:**

❖ Answer **ALL FIVE** questions.

❖ Missing data may be suitably assumed.

---

| | | |
|---|---|---|
| **1A.** | Present a detailed outline of any four aspects of Mining methodology. | **4** |
| **1B.** | Summarize any four architectural types of Data Warehouses. | **4** |
| **1C.** | With the help of sufficient examples examine any 4 descriptors inside the Dimensional table of a Star Schema model | **2** |
| **2A.** | With the help of an example present an elaborate summarization of the general principles and method of application of Type 2 changes to Data Warehouses | **5** |
| **2B.** | Justify how Immediate Data extraction is carried out in Data Warehouses | **3** |
| **2C.** | With the help of an example, compare Interval Scaled and Ratio scaled attribute types | **2** |
| **3A.** | For the given dataset, assuming minimum support is set to a value of 2, generate all frequent itemsets using the FP-Growth algorithm. Show the detailed steps by constructing the Conditional (Sub-)Pattern Bases and also show the conditional FP-tree associated with the conditional node I3 using pictorial representation. <br><br> | **5** |
| **3B.** | Design the Apriori algorithm to discover frequent itemsets for mining Boolean association rules | **3** |
| **3C.** | For the contingency table shown in Table 3C determine what type of correlation exists between buying game and buying video <br><br> Table 3C | **2** |

Table for 3A:

| TID | List of item_IDs |
|---|---|
| T100 | I1, I2, I5 |
| T200 | I2, I4 |
| T300 | I2, I3 |
| T400 | I1, I2, I4 |
| T500 | I1, I3 |
| T600 | I2, I3 |
| T700 | I1, I3 |
| T800 | I1, I2, I3, I5 |
| T900 | I1, I2, I3 |

|  | game | $\overline{game}$ | $\Sigma_{row}$ |
|---|---|---|---|
| video | 4000 | 3500 | 7500 |
| $\overline{video}$ | 2000 | 500 | 2500 |
| $\Sigma_{col}$ | 6000 | 4000 | 10,000 |

| | | |
|---|---|---|
| **4A.** | Consider the following figure showing a multilayer feed-forward neural network. Let the learning rate be 0.9. The initial weight and bias values of the network are given in Table 1, Classify the tuple, $X =(1, 0, 1)$ with a class label of 1 using Backpropagation algorithm. Show all steps in detail for the first iteration.  Initial Input, Weight, and Bias Values | 5 |

| $x_1$ | $x_2$ | $x_3$ | $w_{14}$ | $w_{15}$ | $w_{24}$ | $w_{25}$ | $w_{34}$ | $w_{35}$ | $w_{46}$ | $w_{56}$ | $\theta_4$ | $\theta_5$ | $\theta_6$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0 | 1 | 0.2 | −0.3 | 0.4 | 0.1 | −0.5 | 0.2 | −0.3 | −0.2 | −0.4 | 0.2 | 0.1 |

| | | |
|---|---|---|
| **4B.** | With the help of examples give a detailed outline of the pruning techniques involved in mining Closed and Maximal Itemsets | 3 |
| **4C.** | Present a summary of the commonly used approaches for pruning Decision Trees | 2 |
| **5A.** | With the help of a diagram furnish an elaborate breakdown of the procedure involved in Agglomerative and Divisive Hierarchical clustering of data objects. | 4 |
| **5B.** | Using sufficient illustration present an intricate summary of the working of CHAMELEON algorithm for clustering of data. | 4 |
| **5C.** | With the help of a diagram examine how Support Vector Machines classify data when data are linearly separable | 2 |