B. TECH. (COMPUTER SCIENCE & BUSINESS SYSTEMS) (CBCS - 2018 COURSE)
B.Tech. (CSBS) Sem - VI : WINTER : 2023
SUBJECT : DATA MINING & ANALYTICS

Day : Friday
Date : 24-11-2023

W-20476-2023

Time : 10:00 AM-01:00 PM
Max. Marks : 60

N.B.

1) All questions are **COMPULSORY**.
2) Figures to the right indicate **FULL** marks.
3) Use of non - programmable **CALCULATOR** is allowed.
4) Assume suitable data if necessary.

Q.1 Define and explain data mining with it's stages along with necessary diagram. **(10)**

OR

Q.1 Differentiate between **(10)**
i) OLTP and OLAP
ii) Clustering and classification

Q.2 State and explain the methods of data cleaning in detail. **(10)**

OR

Q.2 Explain data discretization and concept hierarchy generation. **(10)**

Q.3 Discuss the concept of Frequent Pattern Growth algorithm and generate FP tree for the following transaction data set. **(10)**
(Minimum Support = 30%)

Tid	Items
1	E,A,D,B
2	D,A,C,E,B
3	C,A,B,E
4	B,A,D
5	D
6	D,B
7	A,D,E
8	B,C

OR

Q.3 Generate rules using Apriori algorithm for the following given transaction data set. **(10)**

Transaction id	Items
1	Bread, Cheese, Egg, Juice
2	Bread, Cheese, Juice
3	Bread, Milk, Yogurt
4	Bread, Juice, Milk
5	Cheese, Juice, Milk

P.T.O.

Q.4 Explain the following (10)

i) Descriptive and Predictive Analysis

ii) Simple Linear Regression Analysis

OR

Q.4 What are the steps in Hypothesis Testing? Explain with suitable example. (10)

Q.5 Describe the Prediction model types (in the generalized linear model) to fit the following use cases: (10)

i) Number of Website Visitors per Hour

ii) Temperature Vs Pressure readings for an electric furnace

iii) Number of spam emails received by a person.

OR

Q.5 Write in brief the convergence behavior of Newton Raphson's method to minimize the cost function. (10)

Q.6 Describe the following types of time series methods used for forecasting: (10)

i) AR

ii) MA

iii) ARMA

OR

Q.6 What are the types of components of time series plot? Explain in detail (10)

* * *