

B. TECH. (COMPUTER SCIENCE & BUSINESS SYSTEMS) (CBCS - 2018 COURSE)

**B.Tech. (CSBS) Sem - VI :SUMMER- 2022
SUBJECT : DATA MINING & ANALYTICS**

Day : Thursday
Date : 23-06-2022

S-20476-2022

Time : 02:30 PM-05:30 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 **WHEREVER** necessary.
- 5) Draw neat labeled diagrams **wherever** necessary.

- Q.1 a)** Classify the different types of data on which mining can be performed **(05)**
b) Enlist stages of data mining process. Signify the visualization. **(05)**

OR

- Q.1 a)** Discuss about the major issues in data mining. **(05)**
b) Illustrate the architecture of a typical data mining system. **(05)**

- Q.2 a)** How to represent input data and output knowledge? **(05)**
b) Enlist data discretization techniques and explain one with example. **(05)**

OR

- Q.2 a)** With suitable example explain data reduction. **(05)**
b) Compare WEKA and Python (with libraries) as a data mining tools. **(05)**

- Q.3** For the following given transaction data set, generate rules using Apriori Algorithm. Consider the values as support = 22 % and confidence = 70%. **(10)**

Transaction ID	Items purchased
1	11,12,15
2	12,14
3	12,13
4	11,12,14
5	11,13
6	12,13
7	11,13
8	11,12,13,15
9	11,12,13

OR

- Q.3 a)** What is inferring rudimentary rules? Write its algorithm. **(05)**
b) Describe Bayesian classification. **(05)**

- Q.4** Enlist assumptions of linear regression model. In wood cutting shop a piece of wood is cut into two parts. The hourly cutting data is recorded into to following table. Predict 'how much time required to cut 120 wood pieces into 1200 pieces.'

x: No. of hours. y: No. of wood pieces before cutting . xy: No. of wood pieces after cutting

Obs	x	y	xy
1	2	21	42
2	4	27	108
3	6	29	174
4	8	64	512
5	10	86	860
6	12	92	1104

PTO

OR

- Q.4** a) List out methods of hypothesis testing. Explain one in detail. (05)
b) Compare predictive modelling and pattern discovery. (05)

Q.5 Define nonlinear regression. Find two examples from your field of specialization where the regression function of a response variable with one or more predictor variables is a nonlinear function of unknown parameters. (10)

OR

- Q.5** a) Write a note on nonparametric regression models. (05)
b) Enlist semiparametric regression model methods. Illustrate one method. (05)

- Q.6** a) Define autocorrelation. Give its properties & applications. (05)
b) With justification state the correctness of following statement: (05)
“integrated moving average model is better than moving average and autoregressive moving average model”.

OR

- Q.6** a) Explain autoregressive moving average model (ARMA). (05)
b) Define auto-covariance. Give its properties & applications. (05)
