

JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT

TEST -1 EXAMINATION- 2018

B.Tech. (IT), 6th Semester

COURSE CODE: 10B22CI622

MAX. MARKS: 15

COURSE NAME: Data Mining

COURSE CREDITS: 4

MAX. TIME: 1 HR

Note: All questions are compulsory. Carrying of mobile phone during examinations will be treated as case of unfair means.

Q.1. (a) What kind of data can be mined using the data mining? Explain data mining as a step in the process of knowledge discovery. [2 Marks]

(b) Explain how the evolution of database technology led to data mining. [1 Marks]

Q.2. How databases are differs from data ware house. Briefly describe the following *advanced database systems* and applications: object-relational databases, spatial databases, text databases, multimedia databases, the World Wide Web. [3 Marks]

Q.3. Suppose that the data for analysis includes the attribute *age*. The *age* values for the data tuples are (in increasing order for 24 tuples) 13, 20, 21, 22, 23, 23, 24, 24, 25, 25, 25, 25, 30, 30, 31, 32, 33, 35, 35, 35, 36, 40, 52, 70. [3 Marks]

- What is the *mean* of the data? What is the *median*?
- What is the *mode* of the data? Comment on the data's modality (i.e., bimodal, trimodal, etc.).
- What is the *midrange* of the data?
- Can you find (roughly) the first quartile (Q_1) and the third quartile (Q_3) of the data?
- Give the *five-number summary* of the data.
- Show a *boxplot* of the data.

Q.4 Use the two methods below to *normalize* the following group of data: 250; 350; 450; 650; 1100

- min-max normalization by setting $min = 0$ and $max = 1$
- z-score normalization
- Normalization using the decimal method. [3 Marks]

Q.5 Briefly compare the following concepts. You may use an example to explain your point(s).

- Snowflake schema, fact constellation, starnet query model
- Data cleaning, data transformation, refresh
- Enterprise warehouse, data mart, virtual warehouse [3 Marks]