



NATIONAL INSTITUTE OF TECHNOLOGY TIRUCHIRAPPALLI
END SEMESTER EXAMINATION - JAN. 2021 SESSION

DEPARTMENT : Computer Science and Engineering
DATE & TIME OF EXAM : 13-05-2021 & 11.00 a.m. – 01.00 p.m
SUB CODE : CS PE14 DURATION: 2 hours
SUB NAME : Data Warehousing and Data Mining
FACULTY NAME : Dr. E. Sivasankar

Answer all Questions

5 x 10 = 50 marks

1. Suppose a company wants to design a data warehouse to facilitate the analysis of moving vehicles in an online analytical processing manner. The company registers huge amounts of auto movement data in the format of (*Auto ID, location, speed, time*). Each *Auto ID* represents a vehicle associated with information (e.g., *vehicle category, driver category*), and each location may be associated with a street in a city. Assume that a street map is available for the city.
(a) Design a data warehouse schema using galaxy schema to facilitate effective online analytical processing in multidimensional space.
(b) Write the DMQL corresponding to the schema design.
2. Suppose your task as a software engineer at a FMCG company is to design a data mining system. Describe the *architecture* you would choose. What is the purpose of each component of this architecture?
3. Construct a Bayesian classifier model for the following data. Predict the class label for the following attribute values namely Give Birth =yes, Can Fly=no, Live in Water =yes and Have Legs =no.

| Name | Give Birth | Can Fly | Live in Water | Have Legs | Class |
|---------------|------------|---------|---------------|-----------|-------------|
| human | yes | no | no | yes | mammals |
| python | no | no | no | no | non-mammals |
| salmon | no | no | yes | no | non-mammals |
| whale | yes | no | yes | no | mammals |
| frog | no | no | sometimes | yes | non-mammals |
| komodo | no | no | no | yes | non-mammals |
| bat | yes | yes | no | yes | mammals |
| pigeon | no | yes | no | yes | non-mammals |
| cat | yes | no | no | yes | mammals |
| leopard shark | yes | no | yes | no | non-mammals |
| turtle | no | no | sometimes | yes | non-mammals |
| penguin | no | no | sometimes | yes | non-mammals |
| porcupine | yes | no | no | yes | mammals |
| eel | no | no | yes | no | non-mammals |
| salamander | no | no | sometimes | yes | non-mammals |
| gila monster | no | no | no | yes | non-mammals |
| platypus | no | no | no | yes | mammals |
| owl | no | yes | no | yes | non-mammals |
| dolphin | yes | no | yes | no | mammals |
| eagle | no | yes | no | yes | non-mammals |



NATIONAL INSTITUTE OF TECHNOLOGY TIRUCHIRAPPALLI
END SEMESTER EXAMINATION - JAN. 2021 SESSION

4. Considering the following example of a store that sells DVDs, Videos, CDs, Books and Games, the store owner might want to discover which of these items' customers are likely to buy together using Apriori algorithm with min_support=2.

| Transactions | Items |
|--------------|-----------------------|
| Customer1 | BOOKS, CD, VIDEO |
| Customer2 | CD, GAMES |
| Customer3 | CD, DVD |
| Customer4 | BOOKS, CD, GAMES |
| Customer5 | BOOKS, DVD |
| Customer6 | CD, DVD |
| Customer7 | BOOKS, DVD |
| Customer8 | BOOKS, CD, DVD, VIDEO |
| Customer9 | BOOKS, CD, DVD |

5. Briefly explain knowledge discovery in Text databases with focus on statistical techniques used for text summarization.