MCA-II & IV Semester/M.Sc.-II Semester Examinations 2020-2021

Computer Science
Paper: CS303/CS311
(Data Mining)

Time: Four hours]　　　　　　　　　　　　　　　　　　　　　　　[**Full Marks:** 70

Instructions

1. The Question Paper contains 08 questions out of which you are required to answer any 04 questions. The question paper is of 70 marks with each question carrying 17.5 marks.

   प्रश्नपत्र में आठ प्रश्न पूँछे गये हैं जिनमें से 4 प्रश्नों का उत्तर देना है। प्रश्नपत्र 70 अंकों का है, जिसमें प्रत्येक प्रश्न 17.5 अंक का है।

2. The total duration of the examination will be **4 hours** (Four hours), which includes the time for downloading the question paper from the Portal, writing the answers by hand and uploading the hand-written answer sheets on the portal.

   परीक्षा का कुल समय 4 घंटे का है जिसमें प्रश्नपत्र को पोर्टल से डाउनलोड करना, हस्तलिखित प्रश्नों का उत्तर पोर्टल पर अपलोड करना है।

3. For the students with benchmark disability as per Persons with Disability Act, the total duration of examination shall be **6 hours** (six hours) to complete the examination process, which includes the time for downloading the question paper from the Portal, writing the answers by hand and uploading the hand-written answer sheets on the portal.

   दिव्यांग छात्रों के लिये परीक्षा का समय 6 घंटे निर्धारित है जिसमें प्रश्नपत्र को पोर्टल से डाउनलोड करना एवं हस्तलिखित उत्तर को पोर्टल पर अपलोड करना है।

4. Answers should be hand-written on a plain white A4 size paper using black or blue pen. Each question can be answered in upto 350 words on 3 (Three) plain A4 size paper (only one side is to be used).

   हस्तलिखित प्रश्नों का उत्तर एक सादे सफेद A4 साइज के पन्ने पर काले अथवा नीले कलम से लिखा होना चाहिये। प्रत्येक प्रश्न का उत्तर 350 शब्दों अथवा A4 साइज के तीन पृष्ठों का होना चाहिये। प्रश्नों का उत्तर कापी के केवल एक पृष्ठ पर ही लिखना है।

5. Answers to each question should start from a fresh page. All pages are required to be numbered. You should write your Course Name, Semester, Examination Roll Number, Paper Code, Paper title, Date and Time of Examination on the first sheet used for answers.

   प्रत्येक प्रश्न का उत्तर नये पृष्ठ से शुरू करना है। सभी पृष्ठों को पृष्ठांकित करना है। छात्र को प्रथम पृष्ठ पर प्रश्नपत्र का विषय, सेमेस्टर, परीक्षा अनुक्रमांक, प्रश्नपत्र कोड, प्रश्नपत्र का शीर्षक, दिनांक एवं समय लिखना है।

Questions

1. a) Describe the major issues in data mining regarding mining methodology, user interaction, performance, and diverse data types.　(10)

   b) Draw and explain the multi-tier architecture of a data warehouse system.　(7.5)

2. a) With an example, describe all three schema (star/snowflake/fact constellation) definitions using DMQL statements.　(10)

   b) Distinguish between Online Transaction Processing (OLTP) and Online Analytical Processing (OLAP).　(7.5)

3. a) What is data cube? Explain the various data cube operations with examples.　(10)

   b) Briefly explain the various attribute selection measures for classification.　(7.5)

4. a) Define the terms Directly Density-Reachable, Density-Reachable and Density-Connected. Explain DBSCAN algorithm for density based clustering. List out its advantages over K-means.　(10)

   b) Write notes on high performance data mining.　(7.5)

5. a) Discuss the various accuracy and error measures available for model evaluation. (10)

    b) Apply the Apriori algorithm for the following dataset. Use 0.3 for the minimum support value. (7.5)

Trans_ID	Items_Purchased
101	strawberry, litchi, orange
102	strawberry, butter fruit
103	butter fruit, vanilla
104	strawberry, litchi, orange
105	banana, orange
106	banana
107	banana, butter fruit
108	strawberry, litchi, apple, orange
109	apple, vanilla
110	strawberry, litchi

6. a) Explain the k-nearest neighbor classification algorithm with its merits and demerits. (10)

    b) Develop an algorithm for classification using decision tree. Illustrate the algorithm with a relevant example. (7.5)

7. a) Explain the K-Medoids clustering method in detail. Also discuss its advantages over K-Means clustering algorithm. (10)

    b) Write short notes on multimedia data mining. (7.5)

8. a) Describe the applications and trends in data mining in detail. (10)

    b) Discuss the social impacts of data mining in detail. (7.5)