GL BAJAJ
Institute of Management & Research, PGDM Institute
Greater Noida
F I N D   Y O U R   S P A R K

Plot No. 2, Knowledge Park-III, Greater Noida (U.P.) –201306

POST GRADUATE DIPLOMA IN MANAGEMENT (2022-23)
END TERM EXAMINATION (TERM -)

Subject Name: Data mining for Decision making	Time: **02.00 hrs**
Sub. Code: PGIT 51	Max Marks: **40**

INSTRUCTIONS
1. **All questions are to be solved using RapidMiner on individual Computers/LAPTOPS.**
   o **Start with New Process and add new note to the Process : Enter Subject Name, Date, Course & Session (PGDM 2021-23), SEMESTER, Student's Name, ROLL NUMBER on top eg. # Subject Name, # Date and so on**
   o **Use One Process for each question and write # Question number before starting each question**
   o **Export all processes as .rmp files in one folder. Folder must be named with your full name and Roll Number for example (Name_GM-----) . submit the soft copies using a PD. Each student to carry his/ her own Pen Drive (PD).**
   o **Sharing of PD for submitting final ANSWER / SOLUTION is not allowed and would lead to cancellation of the answers submitted.**
   o **Conceptual questions must be answered on answer copy provided by the examination cell and be submitted along with .rmp files in PD**
2. **During examination, no student is allowed to use mobile phones/Smart watch/Internet in any conditions.**
3. **Data sheets (.csv file/Excel) will be provided as a soft copy on the Desktops/Laptops**

All questions are compulsory and carry equal marks   **[8 Marks x 5=40 Marks]**

CO-1 Identify the business needs for knowledge discovery in order to create competitive advantages with data mining technologies appropriately in order to realize their real business value in solving business problems. (L3)
CO-2 Utilize the basic data mining concepts and their application in business context using data mining tools.(L3)
CO-3 Evaluate interesting and useful patterns from the explosive Volume of data by application of supervised and unsupervised techniques. (L3, L4, L5)
CO-4 Examine Integration of theory and application in various functional areas through interdisciplinary approach. (L4)

SECTION - A		
Questions	**CO**	**Bloom's Level**
Q.1 Use the dataset in "**loan_small.csv**" file and do the following: **(A)** Pre-process the dataset to make it clean for further processing in RapidMiner and show the output. **(B)** Use any of the Classification methods to model for loan approval. **(C)** What can be the possible drawbacks of the model ? **(D)** Give another application of Classification method.	CO1, CO2, CO3, CO4	L3,L4,L5
SECTION – B		
Questions	**CO**	**Bloom's Level**
Q. 2: (A). Describe the purpose of data mining through its commercial and		L3

scientific viewpoint.	CO1	
Q. 2: (B). Write the steps of Knowledge discovery and explain why Data pre-processing is important in data mining?		
Q. 3: Use the dataset in "**Country-Data.csv**" file and solve the problem given below using unsupervised learning model. HELP International has been able to raise around $ 10 million. Now the CEO of the NGO needs to decide how to use this money strategically and effectively. So, CEO has to make decision to choose the countries that are in the direct need of aid. Hence, your Job as a Data analyst is to categorize the countries using some socio-economic and health factors that determine the overall development of the country. Then you need to suggest the countries which the CEO needs to focus on the most.	CO2	L3 L3,L4,L5
Q. 4: Use **"Adult-Income.csv"** dataset to predict if a person can make over Rs 50,000 a month, with the help of Decision Tree Model. Also show the systematic flow of operators in RapidMiner.	CO3	

SECTION - C

Questions	CO	Bloom's Level
Q. 5: Bob has started his own mobile company. He wants to give tough fight to big companies like Apple,Samsung etc. He does not know how to estimate price of mobiles his company creates. In this competitive mobile phone market you cannot simply assume things. To solve this problem he collects sales data of mobile phones of various companies. Bob wants to find out some relation between features of a mobile phone(eg:-RAM,Internal Memory etc) and its selling price. But he is not so good at Machine Learning. So he needs your help to solve this problem. Use **"Mobile Data.csv"** dataset	CO4	L4

COs	Marks Allocated
CO1	10
CO2	10
CO3	10
CO4	10