

Roll No.....



Plot No. 2, Knowledge Park-III, Greater Noida (U.P.) –201306  
**POST GRADUATE DIPLOMA IN MANAGEMENT (2023-24)**  
**MID TERM EXAMINATION (TERM -V)**

Subject Name: Data mining for Decision making

Time: **40 mins**

Sub. Code: PGIT 51

Max Marks: **20**

**INSTRUCTIONS**

- All questions are to be solved using RapidMiner on individual Computers/LAPTOPS.**
  - Start with New Process and add new note to the Process : Enter Subject Name, Date, Course & Session (PGDM 2022-24), SEMESTER, Student's Name, ROLL NUMBER on top eg. # Subject Name, # Date and so on
  - Use One Process the question
  - 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 to the invigilator using a PD.
  - Conceptual questions must be answered on Note , within the process , in .rmp file.
- During examination, no student is allowed to use mobile phones/Smart watch/Internet in any condition.**
- Data sheets (.csv file/Excel) will be provided as a soft copy on the Desktops/Laptops**

**One Question with subparts.**

**20 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)

**Case Study**

This dataset provides comprehensive insights into the socio-educational factors(variables) influencing student behavior, with a focus on alcohol consumption and academic performance in high school. Collected through a survey conducted among high school students, the dataset includes various social and gender-related information. It explores the complex interplay between social and educational factors on student behavior, focusing on alcohol consumption patterns.

**Mine the data as:**

Use a supervised learning methods to predict and analyse the student alcohol consumption pattern and identify atleast three most significant predictors(Variables) of the pattern. The data must be processed appropriately to make it usable for the model. What is the performance of the model developed? Write your analysis of the pattern identified also.

## Dataset



Habit survey.csv

## Dataset Features

1. **school:** Student's school (binary: GP (General Public) or MS (Marvel Silver))
2. **gender:** Student's gender (binary: 'F' - female or 'M' - male)
3. **age:** Student age (numeric: 15 to 22)
4. **address:** Type of student's residential address (binary: 'U' - urban or 'R' - rural)
5. **family size:** Family size (binary: 'LE3' - less than or equal to 3 or 'GT3' - greater than 3)
6. **Parent status:** Parents' cohabitation status (binary: 'T' - living together or 'A' - separated)
7. **Mother edu:** Mother's education level (0 - none, 1 - Elementary School 1, 2 - Elementary School 2, 3 - High School or 4 - Higher Education)
8. **Father edu:** Father's education level (0 - none, 1 - Elementary School 1, 2 - Elementary School 2, 3 - High School or 4 - Higher Education)
9. **Mother job:** Mother's job (nominal: teacher, health, services, at\_home or Other)
10. **Father job:** Father's job (nominal: teacher, health, services, at\_home or Other)
11. **School support :** Extra educational support (binary: yes or no)
12. **Family support:** Family educational support (binary: yes or no)
13. **Extracur activities:** Performs extracurricular activities (binary: yes or no)
14. **Daycare:** Attended daycare (binary: yes or no)
15. **Internet access:** Internet access at home (binary: yes or no)
16. **Family rel:** Quality of family relationships (categorical: from 1 - very bad to 5 - excellent)
17. **freetime:** Free time after school (categorical: from 1 - very low to 5 - very high)
18. **goout:** Time with friends (categorical: from 1 - very low to 5 - very high)
19. **Daily alc:** Alcohol consumption on the work day (categorical: from 1 - very low to 5 - very high)
20. **Weekly alc:** Alcohol consumption on the weekend (categorical: from 1 - very low to 5 - very high)
21. **health:** Current health status (categorical: from 1 - very bad to 5 - very good)