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 (2022-24)
END TERM EXAMINATION (TERM -V)

Subject Name: **Data mining for Decision making**	Time: **90 mins**
Sub. Code: **PGIT51**	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 2022-24), SEMESTER, Student's Name, ROLL NUMBER on top eg. # Subject Name, # Date and so on**
   o **Use One Process the 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 to the invigilator using a PD.**
   o **Conceptual questions must be answered on Note , within the process , in .rmp file.**
2. **During examination, no student is allowed to use mobile phones/Smart watch/Internet in any condition.**
3. **Data sheets (.csv file/Excel) will be provided as a soft copy on the Desktops/Laptops**

Two Question with subparts.   **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)

Question 1   **20 marks**

Case Study

Given **Dataset 1**, on Google Playstore Apps, provides comprehensive insights into the preference of the customers, their reviews on installation of various apps. It has 12 factors giving information about the installed apps. It explores some strong interplay between the the rating of the app and some factors. you may help the app providers to focus on these factors which impact rating pattern, through data mining.

Mine the data as:

(a) Use a suitable data mining methods to predict and analyse the app rating pattern and identify atleast two most significant predictors of the pattern.
    a. What factors are responsible for higher app rating by customers?
    b. What category is preferred the most?
    c. What kind of apps should target the teens?
(b) The data must be processed appropriately to make it usable for any data mining method.

(c) Write your analysis of the pattern identified by giving answer to each question.

Question 2   **20 marks**

Case Study

Given Dataset 2, on Google Playstore Apps, provides inclusive perceptions of the customers for the apps they have installed. They have given their review ratings and have some sentiments connected to the service experience. The app providers would like to know the overall review for each type of app. The dataset has just 5 factors but the data is quite dirty. Looking at the potential of the data, it must be analysed to enhance customer base by improving their experience.

Mine the data as:

(a) Use a suitable text mining methods to
   a. Identify the most frequently occurring words?
   b. Identify the major concerns of the customers, if any?
(b) Write all the steps of text mining with suitable description of each and an example.

Dataset Features

Dataset 1

1. **App:** Name of the app (Nominal)
2. **Category:** Category of app (Nominal)
3. **Rating:** Assessment of app quality given by the customers (Ordinal: on the scale of 1(Poor) to 5(Excellent))
4. **Reviews:** Number of reviews on an app (Numeric)
5. **Size:** Size of the app (Numeric)
6. **Installs:** Number of installations of an app(Numeric)
7. **Type:** Category of app (Nominal: Free Or Paid)
8. **Price:** Price of the app if Type is paid (Numeric)
9. **Content Rating:** Type of the content of an app (Nominal: Everyone, Everyone 10+, Teen, Mature 17+, Adults 18 +)
10. **Last Updated:** Updation date of the app(Date)
11. **Current Version:** Latest version of the app(Numeric)
12. **Android Version:** On which version of operating system app can be installed(Numeric)

Dataset 2

13. **App:** Name of the app (Nominal)
14. **Translated_Review:** Review of the customer about the app(Text)
15. **Sentiment:** Customers' feeling (sentiment)
16. **Sentiment_Polarity:** Measure of sentiment (ranges -1 (very negative) to +1(very positive), 0 is neutral)
17. **Sentiment_Subjectivity:** Measure of sentiment(ranges from 0 to 1, 0 means text written based on some evidence/facts, 1 means text written based on personal feeling/opinion)