School of Business

Master of Business Administration MBA Dual Specialization
Mid Term Examination - Mar 2024

Duration : 90 Minutes
Max Marks : 50

Sem IV - MBBA6010 - Data Mining and Predictive Analysis

General Instructions
Answer to the specific question asked
Draw neat, labelled diagrams wherever necessary
Approved data hand books are allowed subject to verification by the Invigilator

1) Compare and contrast supervised and unsupervised classification techqnieus for the given dataset. What are benefits of using each of them. Which is better in your opinion? K5 (5)

Customer ID	Age	Income	Savings	Debt	Credit Score
1	35	50000	20000	1000	600
2	45	75000	30000	5000	700
3	30	40000	15000	2000	550
4	50	80000	40000	8000	750
...	...	...	...	...	...

2) A government agency is dedicated to bolstering disaster preparedness for decision-making to identifyng the likelihood of future natural disasters in a specific region. How can it use the data mining of historical disaster data for its goal? K3 (6)

3) Retail industries utilise classifiers to make specific business decisions. Compare the impact of different distance metrics in nearest neighbors algorithms. How does the choice of distance metric influence the clustering or classification outcomes? K4 (8)

4) Amazon.com uses association rules to display "Frequently bought together" items under product listings. This helps customers to discover complementary products and increases the likelihood of additional purchases. Demonstrate the benefits of using collaborative filtering and its uses in recommendation systems for the given dataset. K3 (9)

User ID	Product ID	Product Name	Category	Quantity
101	1001	Digital Camera	Electronics	1
101	1002	Camera Bag	Accessories	1
102	1003	Smartphone	Electronics	1
102	1004	Bluetooth Speaker	Electronics	1
103	1005	Headphones	Electronics	1
103	1006	Power Bank	Accessories	1
...	...	...	...	...

5) An e-commerce platform wants to classify customers based on their purchase behavior to tailor marketing strategies and promotions more effectively. The goal is to identify customers who make frequent purchases versus those who make infrequent purchases, allowing the company to personalize marketing efforts and increase customer engagement. This company has hired you to implement data mining for better analysis and insights. Compare and contrast supervised and non supervised data miningalgorithms that could be utilsed.

6) Design an apporach using knn algorithm to classify appraisal class using given input variables such as with the help of age, distance from home, Education, hourly rate, monthly income, percent salary hike, total working years, years in current role, years since last promotion, years with current manager, Gender and marital status etc. The snippet of the dataset is given below. What would be the expected output of this model? Could you suggest any alternative model?

Employee ID	Age	Distance from Home	Education	Hourly Rate	Monthly Income	Percent Salary Hike	Total Working Years	Years in Current Role	Years Since Last Promotion	Years with Current Manager	Gender	Marital Status	Appraisal Class
1	30	5	3	25	5000	10	7	4	2	3	Male	Single	Good
2	35	10	4	30	6000	15	10	7	4	5	Female	Married	Excellent
3	40	3	5	35	7000	12	15	10	5	10	Male	Single	Average
4	45	15	3	28	5500	8	20	15	8	12	Female	Married	Good
5	28	2	2	20	4000	20	5	2	1	1	Male	Single	Poor