GALGOTIAS UNIVERSITY

ADMISSION NUMBER

School of Business

Master of Business Administration MBA Dual Specialization
Semester End Examination - Aug 2024

Duration : 180 Minutes
Max Marks : 100

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)	Identify the strengths and weaknesses of sentiment analysis techniques in processing customer feedback for improving service quality and driver performance in the ride-sharing industry, considering implications for customer satisfaction and brand reputation.	K3(6)
2)	A retail company has provided you a data set to aid them in analsysis. Analyse the given dataset for a hypothetical scenario where TP is 40, FP is 20, FN is 15, and TN is 130. Compute precision and recall for this scenario, then interpret the obtained results. (TP - True Positive, FP - False Positive, FN - False Negative, TN - True Negative)	K4(8)
3)	Analyse precision and recall in the context of medical diagnosis and the importance of each parameter.	K4(8)
4)	You are being provided with the following dataset of grocery items. Dissect the data and provide any two association rules from the given transaction dataset.	K4(4)

Transaction ID	Items Purchased
1	Bread, Milk, Eggs
2	Bread, Milk, Cheese
3	Milk, Cheese
4	Bread, Eggs, Cheese
5	Bread, Milk
6	Bread, Cheese
7	Milk, Eggs
8	Bread, Milk, Eggs
9	Bread, Cheese
10	Milk, Eggs

5)	How does a local dairy store apply the concept of support to enhance sales of butter? What is the value of support for product	K3(9)

butter?

Transaction ID	Customer ID	Product ID	Product Name	Quantity
1	101	1001	Butter	1
1	101	1002	Bread	2
2	102	1001	Butter	1
2	102	1003	Milk	1
3	103	1002	Bread	1
3	103	1004	Cheese	1
4	104	1001	Butter	1
4	104	1005	Eggs	1
...	...	...	...	...

6) An established e-commerce company is gearing up for extending its product range as part of its strategy for scaling up and global expansion. Compare the use of association rule mining and collaborative filtering for recommending products to customers. K5(10)

7) As part of your research on medication efficacy, you've collected and performed analysis as per the following dataset. Evaluate how hospital administrators can assess the effectiveness of different treatment options for patients with a specific medical condition based on the outcomes of success or failure, and present an analysis suitable for submission to hospital administration. K5(10)

Treatment Option	Successful Patients (Observed)	Unsuccessful Patients (Observed)	Total Patients (Observed)	Successful Patients (Expected)	Unsuccessful Patients (Expected)	Total Patients (Expected)	Chi-square Value
Medication A	15	10	25	20	5	25	1.25
Medication B	12	8	20	15	5	20	0.6
Surgery	18	7	25	15	10	25	0.6
Physical Therapy	20	5	25	20	5	25	0
New Treatment 1	10	5	15	12	3	15	0.333
New Treatment 2	8	12	20	10	10	20	0.4
Total	83	47	130	92	38	130	3.183

Null Hypothesis	Critical Value	Calculated Chi-square Value	Conclusion
Independence	9.4889.488	3.1833.183	Fail to Reject Null Hypothesis

8) In a case study centered on the Indian retail system, explain how the strategic integration of data mining techniques could enhance operational efficiency. Additionally, elucidate the critical aspects that should be thoroughly evaluated to determine the feasibility and assess the impact of these techniques on optimizing resources within the system. K5(15)

9) You're analyzing customer segmentation for a retail chain based upon the following dataset. Develop a clustering model using the k-means algorithm to identify distinct customer segments based on purchasing behavior and demographic information. Discuss the model's performance using metrics like Silhouette Score and discuss how the identified segments can inform marketing strategies. K6(18)

Customer ID	Age	Gender	Annual Income (USD)	Total Purchases ($)
1	30	Male	40000	500
2	45	Female	60000	800
3	25	Female	35000	400
4	50	Male	70000	1000
5	35	Female	50000	600
6	55	Male	80000	1200
7	28	Male	30000	300
8	40	Female	55000	700
9	48	Male	75000	900
10	32	Female	45000	550

Cluster	Average Age	Average Annual Income (USD)	Average Total Purchases (USD)	Number of Customers
0	51.00	76,666.67	1,050.00	3
1	28.67	31,666.67	366.67	3
2	45.00	60,000.00	800.00	4

10) A fintech startup aims to evaluate the effectiveness of its fraud detection system in detecting fraudulent transactions as given in the following table. Discuss the performance of supervised learning techniques in fraud detection and discuss considerations for model deployment in financial institutions. K6(12)

Transaction Amount	Merchant Category	Time of Transaction	Previous Transactions	Fraudulent
1000	Retail	08:30	3	No
1500	Online Shopping	12:45	5	No
200	Restaurant	18:20	2	Yes
3000	Electronics	09:10	4	No
500	Grocery	15:00	6	No
250	Retail	17:30	1	Yes
700	Online Shopping	11:20	7	No
1200	Electronics	14:15	2	No
180	Restaurant	16:40	3	Yes
2200	Electronics	10:00	8	No

True Positives (TP): 2 , False Positives (FP): 1 , True Negatives (TN): 5 , False Negatives (FN): 2