



**SILVER OAK
UNIVERSITY**
EDUCATION TO INNOVATION

Silver Oak College of Engineering and Technology
College of Technology (01)
Bachelor of Technology
Department of Information Technology(010)

Semester:	VI	Academic Year:	2023-24
Subject Name:	DMBI and DMDW	Subject Code:	1010103324/ 1010103323

QUESTION BANK

Sr. No.	Question	CO	Bloom's Taxonomy	Marks
UNIT: 1 OVERVIEW AND CONCEPTS OF DATA WAREHOUSING AND BUSINESS INTELLIGENCE				
1	Define following terms & differentiate them: Data Mart , Enterprise Warehouse & Virtual Warehouse, Closed frequent item set, Outlier Analysis	CO1	U	06
2	Do feature wise comparison between BI and DW	CO1	U	04
3	Explain various features of Data Warehouse?	CO1	U	03
4	Discuss the application of data warehousing and data mining.	CO1	R	04
5	Explain Star, Snowflake, and Fact Constellation Schema for Multidimensional Database.	CO1	U	06
6	Give differences between OLAP and OLTP	CO1	R	04
7	Explain various OLAP operations.	CO1	U	06
8	Explain three tier data warehouse Architecture in details.	CO1	U	06
9	Compare data mart and data warehouse.	CO1	R	03

UNIT: 2 INTRODUCTION TO DATA MINING AND DATA PRE-PROCESSING

9	Define the term “data mining”. Discuss the major issues in data mining.	CO2	R	03
10	What is Data Mining? Why is it called data mining rather knowledge mining? Explain KDD process.	CO2	U	06
11	What is noise? Explain data smoothing methods as noise removal technique to divide given data into bins of size 3 by bin partition (equal frequency), by bin means, by bin medians and by bin boundaries. Consider the data:10, 2, 19, 18, 20, 18, 25, 28, 22	CO2	AN	04
12	Minimum salary is 20,000Rs and Maximum salary is 1,70,000 Rs. Map the salary 1,00,000 Rs in new Range of (60,000 , 2,60,000) Rs using min-max normalization method.	CO2	AN	04
13	If Mean salary is 54,000Rs and standard deviation is 16,000 Rs then find z-score value of 73,600 Rs salary.	CO2	AN	03
14	Explain Mean, Median, Mode Variance, Standard Deviation & five number summaries with suitable database example.	CO2	U	04
15	Enlist data reduction strategies and explain any two.	CO2	U	04
16	Explain the following terms: Numerosity reduction, Data Integration, Data transformation	CO2	R	03
17	Enlist the preprocessing steps with example. Explain procedure of any technique of preprocessing.	CO2	U	06

UNIT: 3 DATA MINING TECHNIQUES

18	What is market basket analysis? Explain the two measures of rule interestingness: support and confidence with suitable example.	CO2	U	04
19	Consider following database of ten transactions. Let min_sup = 30% and min_confidence = 60%. A) Find all frequent itemsets using Apriori algorithm. B) Generate strong association rules	CO2	U/A	06

	<table><tr><th>TID</th><th>items bought</th></tr><tr><td>T1</td><td>pen, pencil</td></tr><tr><td>T2</td><td>book, eraser, pencil</td></tr><tr><td>T3</td><td>book, chalk, eraser, pen</td></tr><tr><td>T4</td><td>chalk, eraser, pen</td></tr><tr><td>T5</td><td>book, pen, pencil</td></tr><tr><td>T6</td><td>book, eraser, pen, pencil</td></tr><tr><td>T7</td><td>ink, pen</td></tr><tr><td>T8</td><td>book, pen, pencil</td></tr><tr><td>T9</td><td>eraser, pen, pencil</td></tr><tr><td>T10</td><td>book, chalk, pencil</td></tr></table>	TID	items bought	T1	pen, pencil	T2	book, eraser, pencil	T3	book, chalk, eraser, pen	T4	chalk, eraser, pen	T5	book, pen, pencil	T6	book, eraser, pen, pencil	T7	ink, pen	T8	book, pen, pencil	T9	eraser, pen, pencil	T10	book, chalk, pencil			
TID	items bought																									
T1	pen, pencil																									
T2	book, eraser, pencil																									
T3	book, chalk, eraser, pen																									
T4	chalk, eraser, pen																									
T5	book, pen, pencil																									
T6	book, eraser, pen, pencil																									
T7	ink, pen																									
T8	book, pen, pencil																									
T9	eraser, pen, pencil																									
T10	book, chalk, pencil																									
20	What are the limitations of the Apriori approach for mining? Briefly describe the techniques to improve the efficiency of Apriori algorithm with an example.	CO2	U/A	06																						
21	<p>Generate frequent item sets and generate association rules based on it using Apriori algorithm. Minimum support is 60% and minimum confidence is 70%</p> <table><tr><th>TID</th><th>Items</th></tr><tr><td>100</td><td>1, 3, 4</td></tr><tr><td>200</td><td>2, 3, 5</td></tr><tr><td>300</td><td>1, 2, 3, 5</td></tr><tr><td>400</td><td>2, 5</td></tr></table>	TID	Items	100	1, 3, 4	200	2, 3, 5	300	1, 2, 3, 5	400	2, 5	CO2	U/A	06												
TID	Items																									
100	1, 3, 4																									
200	2, 3, 5																									
300	1, 2, 3, 5																									
400	2, 5																									

22	<p>State the Apriori Property. Generate large itemsets and association rules using Apriori algorithm on the following data set with minimum support value and minimum confidence value set as 50% and 75% respectively.</p> <table><tr><th><i>TID</i></th><th><i>Items Purchased</i></th></tr><tr><td>T101</td><td>Cheese, Milk, Cookies</td></tr><tr><td>T102</td><td>Butter, Milk, Bread</td></tr><tr><td>T103</td><td>Cheese, Butter, Milk, Bread</td></tr><tr><td>T104</td><td>Butter, Bread</td></tr></table>	<i>TID</i>	<i>Items Purchased</i>	T101	Cheese, Milk, Cookies	T102	Butter, Milk, Bread	T103	Cheese, Butter, Milk, Bread	T104	Butter, Bread	C02	U/A	06
<i>TID</i>	<i>Items Purchased</i>													
T101	Cheese, Milk, Cookies													
T102	Butter, Milk, Bread													
T103	Cheese, Butter, Milk, Bread													
T104	Butter, Bread													

UNIT: 4 CLASSIFICATION AND PREDICTION

23	What is classification and prediction? List out Issues regarding Classification and prediction	C03	R/U	04																																																																																										
24	Explain the steps of the ID3 algorithm for generating Decision trees.	C03	A	06																																																																																										
25	Why naïve Bayesian classification is called “naïve”? Describe naïve Bayesian classification with example.	C03	A	06																																																																																										
26	Do feature wise comparison between classification and prediction	C03	U	04																																																																																										
27	Generate decision tree using ID3 algorithm for the following dataset. <table><tr><th>Sr. no.</th><th>Outlook</th><th>Temperature</th><th>Humidity</th><th>Wind</th><th>Play</th></tr><tr><td>1</td><td>Sunny</td><td>hot</td><td>high</td><td>FALSE</td><td>No</td></tr><tr><td>2</td><td>Sunny</td><td>hot</td><td>high</td><td>TRUE</td><td>No</td></tr><tr><td>3</td><td>Overcast</td><td>hot</td><td>high</td><td>FALSE</td><td>Yes</td></tr><tr><td>4</td><td>Rain</td><td>mild</td><td>high</td><td>FALSE</td><td>Yes</td></tr><tr><td>5</td><td>Rain</td><td>cool</td><td>normal</td><td>FALSE</td><td>Yes</td></tr><tr><td>6</td><td>Rain</td><td>cool</td><td>normal</td><td>TRUE</td><td>No</td></tr><tr><td>7</td><td>Overcast</td><td>cool</td><td>normal</td><td>TRUE</td><td>Yes</td></tr><tr><td>8</td><td>Sunny</td><td>mild</td><td>high</td><td>FALSE</td><td>No</td></tr><tr><td>9</td><td>Sunny</td><td>cool</td><td>normal</td><td>FALSE</td><td>Yes</td></tr><tr><td>10</td><td>Rain</td><td>mild</td><td>normal</td><td>FALSE</td><td>Yes</td></tr><tr><td>11</td><td>Sunny</td><td>mild</td><td>normal</td><td>TRUE</td><td>Yes</td></tr><tr><td>12</td><td>Overcast</td><td>mild</td><td>high</td><td>TRUE</td><td>Yes</td></tr><tr><td>13</td><td>Overcast</td><td>hot</td><td>normal</td><td>FALSE</td><td>Yes</td></tr><tr><td>14</td><td>Rain</td><td>mild</td><td>high</td><td>TRUE</td><td>No</td></tr></table>	Sr. no.	Outlook	Temperature	Humidity	Wind	Play	1	Sunny	hot	high	FALSE	No	2	Sunny	hot	high	TRUE	No	3	Overcast	hot	high	FALSE	Yes	4	Rain	mild	high	FALSE	Yes	5	Rain	cool	normal	FALSE	Yes	6	Rain	cool	normal	TRUE	No	7	Overcast	cool	normal	TRUE	Yes	8	Sunny	mild	high	FALSE	No	9	Sunny	cool	normal	FALSE	Yes	10	Rain	mild	normal	FALSE	Yes	11	Sunny	mild	normal	TRUE	Yes	12	Overcast	mild	high	TRUE	Yes	13	Overcast	hot	normal	FALSE	Yes	14	Rain	mild	high	TRUE	No	C03	AN	06
Sr. no.	Outlook	Temperature	Humidity	Wind	Play																																																																																									
1	Sunny	hot	high	FALSE	No																																																																																									
2	Sunny	hot	high	TRUE	No																																																																																									
3	Overcast	hot	high	FALSE	Yes																																																																																									
4	Rain	mild	high	FALSE	Yes																																																																																									
5	Rain	cool	normal	FALSE	Yes																																																																																									
6	Rain	cool	normal	TRUE	No																																																																																									
7	Overcast	cool	normal	TRUE	Yes																																																																																									
8	Sunny	mild	high	FALSE	No																																																																																									
9	Sunny	cool	normal	FALSE	Yes																																																																																									
10	Rain	mild	normal	FALSE	Yes																																																																																									
11	Sunny	mild	normal	TRUE	Yes																																																																																									
12	Overcast	mild	high	TRUE	Yes																																																																																									
13	Overcast	hot	normal	FALSE	Yes																																																																																									
14	Rain	mild	high	TRUE	No																																																																																									

28	Explain Logistic Regression with an example.	CO3	U	04																																																																											
29	Explain Linear regression with example.	CO3	U	04																																																																											
30	<p>Using Naive Bayesian classification method, predict class label of X = (age = youth, income = medium, student = yes, credit_rating = fair) using following training dataset.</p> <table border="1"> <thead> <tr> <th>age</th><th>income</th><th>Student</th><th>credit_rating</th><th>Class: buys_computer</th></tr> </thead> <tbody> <tr><td>youth</td><td>high</td><td>no</td><td>Fair</td><td>no</td></tr> <tr><td>youth</td><td>high</td><td>no</td><td>excellent</td><td>no</td></tr> <tr><td>middle_aged</td><td>high</td><td>no</td><td>fair</td><td>yes</td></tr> <tr><td>senior</td><td>medium</td><td>no</td><td>fair</td><td>yes</td></tr> <tr><td>senior</td><td>low</td><td>yes</td><td>fair</td><td>yes</td></tr> <tr><td>senior</td><td>low</td><td>yes</td><td>excellent</td><td>no</td></tr> <tr><td>middle_aged</td><td>low</td><td>yes</td><td>excellent</td><td>Yes</td></tr> <tr><td>youth</td><td>medium</td><td>no</td><td>fair</td><td>no</td></tr> <tr><td>youth</td><td>low</td><td>yes</td><td>fair</td><td>yes</td></tr> <tr><td>senior</td><td>medium</td><td>yes</td><td>fair</td><td>yes</td></tr> <tr><td>youth</td><td>medium</td><td>yes</td><td>excellent</td><td>yes</td></tr> <tr><td>middle_aged</td><td>medium</td><td>no</td><td>excellent</td><td>yes</td></tr> <tr><td>middle_aged</td><td>high</td><td>yes</td><td>fair</td><td>yes</td></tr> <tr><td>senior</td><td>medium</td><td>no</td><td>excellent</td><td>no</td></tr> </tbody> </table>	age	income	Student	credit_rating	Class: buys_computer	youth	high	no	Fair	no	youth	high	no	excellent	no	middle_aged	high	no	fair	yes	senior	medium	no	fair	yes	senior	low	yes	fair	yes	senior	low	yes	excellent	no	middle_aged	low	yes	excellent	Yes	youth	medium	no	fair	no	youth	low	yes	fair	yes	senior	medium	yes	fair	yes	youth	medium	yes	excellent	yes	middle_aged	medium	no	excellent	yes	middle_aged	high	yes	fair	yes	senior	medium	no	excellent	no	CO3	AN	06
age	income	Student	credit_rating	Class: buys_computer																																																																											
youth	high	no	Fair	no																																																																											
youth	high	no	excellent	no																																																																											
middle_aged	high	no	fair	yes																																																																											
senior	medium	no	fair	yes																																																																											
senior	low	yes	fair	yes																																																																											
senior	low	yes	excellent	no																																																																											
middle_aged	low	yes	excellent	Yes																																																																											
youth	medium	no	fair	no																																																																											
youth	low	yes	fair	yes																																																																											
senior	medium	yes	fair	yes																																																																											
youth	medium	yes	excellent	yes																																																																											
middle_aged	medium	no	excellent	yes																																																																											
middle_aged	high	yes	fair	yes																																																																											
senior	medium	no	excellent	no																																																																											
31	Explain the Random Forest Algorithm with an example.	CO3	U	06																																																																											
32	Explain the Naive Bayes Classifier with an example.	CO3	U	06																																																																											
33	What is classification? Explain classification as a two-step process with diagram.	CO3	U	04																																																																											
34	Briefly explain Linear and Non-linear regression.	CO3	U	04																																																																											

UNIT: 5 DATA MINING FOR BUSINESS INTELLIGENCE APPLICATIONS

35	Briefly explain the life-cycle of Data Analytics and discuss the role of data scientists.	CO3	U	06
36	Discuss applications of data mining in Banking and Finance.	CO3	U	04
37	How data Mining is useful for Business Intelligence applications viz. Balanced Scorecard, Fraud Detection, Clickstream Mining, Market Segmentation, retail industry, telecommunications industry, banking & finance and CRM	CO3	AN	06
38	Discuss fraud detection and click-stream analysis using data mining.	CO3	U	06
39	Explain data mining application for fraud detection.	CO3	U	03

UNIT: 6 R PROGRAMMING FOR DATA ANALYSIS

40	Define R programming language. Discuss its significance in data analysis.	CO4	R	04
41	Explain the purpose of the help() function in R. How is it used?	CO4	A	03
42	Define vectors in R. Provide examples of creating and manipulating vectors.	CO4	R	04
43	Discuss the significance of factors in R. How are they useful in data analysis?	CO4	U	03
44	Discuss the significance of the “ filter() ” function in “ dplyr ”. How is it used to subset data?	CO4	U/A	06
45	What are lists in R? Provide an example demonstrating their use.	CO4	R	06
46	Discuss the significance of dimensions in arrays and matrices in R.	CO4	U	04
47	Describe methods for imputing missing values in R. Provide examples using relevant functions or packages.	CO4	R	06
48	Describe the purpose of the ggplot2 package in R. Provide examples of creating different types of plots.	CO4	R	06

49	Discuss the difference between geom_point() and geom_line() in ggplot2. When would you use each?	CO4	U/A	06
50	Explain the process of fitting a linear regression model in R using the lm() function. Provide a code example.	CO4	U/A	06
51	Explain how to visualize time series data in R using plots such as time series plots and seasonal decomposition plots.	CO4	R	06
UNIT: 7 TABLEAU FOR BUSINESS INTELLIGENCE VISUALIZATION				
52	Describe the main components of Tableau Desktop and their roles in the data visualization process.	CO5	R	04
53	How do you connect Tableau Desktop to various data sources? Provide examples.	CO5	U/A	06
54	Create basic visualizations such as bar charts, line charts, and scatter plots in Tableau. Provide step-by-step instructions.	CO5	U/A	03
55	Explain how Tableau supports data exploration through features like drill-down, filtering, and highlighting.	CO5	U	04
56	What is the purpose of Tableau Prep and how does it complement Tableau Desktop for data preparation and cleaning?	CO5	R	06
57	Describe techniques for implementing dynamic and parameterized visualizations in Tableau.	CO5	R	04
58	Discuss the use of custom SQL queries and R/Python integration for advanced data analysis in Tableau.	CO5	U/A	06
59	Describe techniques for optimizing Tableau data extracts for improved performance and scalability.	CO5	R	04
UNIT: 8 ADVANCED TOPICS				
60	Describe the key components of Power BI Desktop.	CO5	R	04
61	Explain the purpose of the Query Editor in Power BI Desktop. How is it used for data transformation?	CO5	R	04
62	Explain the architecture of PowerBI in detail.	CO5	U	06

63	Compare main features and capabilities of Power BI and Tableau.	C05	R	06
64	Describe the process of publishing Power BI reports to the Power BI Service.	C05	R	04
65	Explain the concept of data modeling in Power BI. How does it differ from data transformation?	C05	R	04
66	Explain web mining using example.	C05	R	03
67	Explain Spatial mining using example.	C05	R	03
68	Explain text mining using example.	C05	R	03

Prof. Parvin Bloch
Name & Sign
Subject Co-ordinator

Dr. Rahul Vaghela
Name & Sign
Head of Department