



Dr. Vishwanath Karad

**MIT WORLD PEACE
UNIVERSITY PUNE**

TECHNOLOGY, RESEARCH, SOCIAL INNOVATION & PARTNERSHIPS

School of Computer science and Engineering

Department of Computer Engineering and Technology

SY BTech (AI DS) (Academic Year 2024-25)

Mid Term Exam – Semester IV

Course Name:- Data Engineering Concepts

Course Code:-AID2PR03A

Maximum Marks: 30

Time: 1 Hr

Date:04/03/2025

Instructions:-

1. Attempt any 3 questions from Q. 1 to Q. 4 AND Attempt any 3 questions from Q. 5 to Q. 8
2. Figure to the right indicates full marks.
3. Use of cell phone is prohibited in the examination hall.
4. Neat diagrams must be drawn wherever necessary.
5. Assume suitable data, if necessary and clearly state.
6. Use of scientific calculator is allowed

Attempt any 3 questions from Q. 1 to Q. 4			
Q.1	CO1 Analyzing	Differentiate between structured and unstructured data.	[5 Marks]
Q.2	CO1 Analyzing	Illustrate Knowledge Discovery in Database (KDD) process with the help of a neat diagram.	[5 Marks]
Q.3	CO1 Analyzing	Compare between mean and median. (2) Find mean, median and mode for the following data: 5, 4, 2, 3, 2, 1, 5, 4, 5 (3)	[5 Marks]
Q.4	CO1 Understanding	Explain in detail time series data and geographical data.	[5 Marks]
Attempt any 3 questions from Q. 5 to Q. 8			
Q.5	CO2 Understanding	Why data pre-processing is important? Explain any one method of data pre-processing.	[5 Marks]

Roll No. _____

Q.6	CO2 Applying	Suppose a group of 12 sales price records has been sorted as follows: 5, 10, 11, 13, 15, 35, 50, 55, 72, 92, 204, 215 Partition them into three bins. Solve it by each of the following methods: (a) Equi-depth partitioning (b) Smoothing by bin boundaries	[5 Marks]
Q.7	CO2 Evaluating	Justify the importance of data integration. Explain the issues that need to be addressed during data integration.	[5 Marks]
Q.8	CO2 Understand ing	Explain how outlier analysis and regression are used for data cleaning.	[5 Marks]

04/06/25



Dr. Vishwanath Karad
**MIT WORLD PEACE
 UNIVERSITY** | PUNE
TECHNOLOGY · HUMANITY · SOCIAL INNOVATION · A FUTURE FOR ALL



PRN:

Term End Examination

May/June 2025

AID2PR03A/CET2012B - Data Engineering Concepts

Question Paper ID: 056671

Faculty/School	School of Computer Science and Engineering	Term	Semester IV
Program	SY B.Tech AIDS	Duration	1 hour 30 minutes
Specialization	-	Max. Marks	40

Section - 1 (8 X 5 Marks)

Answer **any 8** questions

1	Explain why data engineering is important in modern data ecosystems.	5 marks	CO1	Understanding														
2	Explain why data cleaning is crucial before analysis. Illustrate with examples.	5 marks	CO1, CO2	Remembering														
3	Compare and contrast the star and snowflake schema in terms of storage and query performance.	5 marks	CO3	Applying														
4	Apply the concept of fact constellation to design a schema for an e-commerce system.	5 marks	CO3, CO4	Applying														
5	Analyze the three-tier architecture of a data warehouse. What are the roles of each tier, and how does this structure improve performance and scalability?	5 marks	CO3, CO4	Analysing														
6	A movie streaming service wants to analyze which genres are frequently watched together. From 6 user sessions, we have the following data. Use the Apriori algorithm with: • Minimum support = 0.5 (50%) • Minimum confidence = 0.7 (70%) • Generate association rules from the frequent itemsets that meet the confidence threshold. <table><tr><th>Transaction ID</th><th>Genres Watched</th></tr><tr><td>T1</td><td>Action, Adventure, Sci-Fi</td></tr><tr><td>T2</td><td>Action, Adventure</td></tr><tr><td>T3</td><td>Action, Sci-Fi</td></tr><tr><td>T4</td><td>Comedy, Romance</td></tr><tr><td>T5</td><td>Action, Adventure, Comedy</td></tr><tr><td>T6</td><td>Adventure, Sci-Fi</td></tr></table>	Transaction ID	Genres Watched	T1	Action, Adventure, Sci-Fi	T2	Action, Adventure	T3	Action, Sci-Fi	T4	Comedy, Romance	T5	Action, Adventure, Comedy	T6	Adventure, Sci-Fi	5 marks	CO4, CO5	Analysing
Transaction ID	Genres Watched																	
T1	Action, Adventure, Sci-Fi																	
T2	Action, Adventure																	
T3	Action, Sci-Fi																	
T4	Comedy, Romance																	
T5	Action, Adventure, Comedy																	
T6	Adventure, Sci-Fi																	

7	A small cafe wants to find frequent combinations of items purchased together. They collected the following 6 transactions. Use the FP-Growth algorithm to find frequent itemsets with minimum support = 0.5 (50%).	5 marks	CO4, CO5	Evaluating														
<table><tr><th>Transaction ID</th><th>Items Bought</th></tr><tr><td>T1</td><td>Milk, Bread, Butter</td></tr><tr><td>T2</td><td>Bread, Butter</td></tr><tr><td>T3</td><td>Milk, Bread, Butter, Coffee</td></tr><tr><td>T4</td><td>Bread, Butter, Jam</td></tr><tr><td>T5</td><td>Milk, Bread, Butter</td></tr><tr><td>T6</td><td>Milk, Bread</td></tr></table>		Transaction ID	Items Bought	T1	Milk, Bread, Butter	T2	Bread, Butter	T3	Milk, Bread, Butter, Coffee	T4	Bread, Butter, Jam	T5	Milk, Bread, Butter	T6	Milk, Bread			
Transaction ID	Items Bought																	
T1	Milk, Bread, Butter																	
T2	Bread, Butter																	
T3	Milk, Bread, Butter, Coffee																	
T4	Bread, Butter, Jam																	
T5	Milk, Bread, Butter																	
T6	Milk, Bread																	
8	Design a Market Basket Analysis system for a grocery store using either Apriori or FP-Growth.	5 marks	CO4, CO5	Creating														
9	Differentiate between supervised and unsupervised learning with suitable real-world examples.	5 marks	CO4	Analysing														
10	Design a classification model using decision trees for student performance prediction.	5 marks	CO4	Creating														

END OF QUESTION PAPER