

<b>Name:</b>	 <b>UPES</b> <small>UNIVERSITY WITH A PURPOSE</small>
<b>Enrolment No:</b>	

**UNIVERSITY OF PETROLEUM AND ENERGY STUDIES**  
**End Semester Examination, May 2023**

<b>Course: Overview of Data Mining</b> <b>Program: MBA-BA KPMG</b> <b>Course code: DSBA-7011</b>	<b>Semester: II</b> <b>Time: 03 hrs.</b> <b>Max. Marks: 100</b>
--------------------------------------------------------------------------------------------------------	-----------------------------------------------------------------------

**SECTION A**  
**10Qx2M= 20 Marks**

- 1. Each Question will carry 2 Marks**  
**2. Instruction: Select/Write the correct answer(s)**

S. No.	Question	CO
Q1.	<p>1. Which of the following is not a open source data mining tool.</p> <p>A. WEKA  B. R  C. RapidMiner  D. KnowledgeMiner</p> <p>2. Classification rules are extracted from_____.</p> <p>A. root node.  B. decision tree.  C. siblings.  D. branches.</p> <p>3. What is the difference between traditional programming and machine learning?</p> <p>a) Traditional programming relies on predefined rules, while machine learning relies on data patterns.  b) Traditional programming requires large amounts of data, while machine learning does not.  c) Traditional programming is faster than machine learning.  d) Traditional programming does not require programming skills, while machine learning does.</p> <p>4. _____analysis divides data into groups that are meaningful, useful, or both.</p> <p>A. Cluster.  B. Association.  C. Classification.  D. Relation.</p> <p>5. The algorithms that are controlled by human during their execution is _____ algorithm.</p> <p>A. unsupervised.</p>	<b>CO1</b>

	<p>B. supervised. C. batch learning. D. incremental.</p> <p>6. Data mining algorithms require _____ A. efficient sampling method. B. storage of intermediate results. C. capacity to handle large amounts of data. D. All of the above.</p> <p>7. In K-nearest neighbor algorithm K stands for _____. A. number of neighbors that are investigated. B. number of iterations. C. number of total records. D. random number.</p> <p>8. Hidden knowledge can be found by using _____. A. searching algorithm. B. pattern recognition algorithm. C. searching algorithm. D. clues.</p> <p>9. Which of the following tasks can be performed using Weka? a. Data preprocessing b. Classification c. Clustering d. All of the above</p> <p>10. What is unsupervised learning? a) The algorithm learns from labeled data b) The algorithm learns from unlabeled data c) The algorithm learns from feedback in a dynamic environment d) The machine learning algorithm is not given any data to train on.</p>	
--	----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--

**SECTION B**  
**4Qx5M= 20 Marks**

- 1. Each question will carry 5 marks**  
**2. Instruction: Write short/brief notes**

Q 2.	<p><b>Differentiate between the following:</b></p> <p>a) Supervised and Unsupervised learning b) AI v/s ML v/s DL c) Classification &amp; Clustering d) Training data and Test data</p>	<b>CO2</b>
------	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	------------

**SECTION-C**  
**3Qx10M=30 Marks**

- 1. Each Question carries 10 Marks.**  
**2. Instruction: Write a long answer.**

Q1.	Explain the following:- i. K-Mean Clustering algorithm. ii. K-Nearest Neighbourhood algorithm	<b>CO2</b>
Q2.	What types of Machine Learning Algorithms exist based on Supervision and how can you illustrate each type using a relevant example?	<b>CO2</b>
Q3.	How can data mining benefit business analyst and what are some common areas of application for this technology?	<b>CO2</b>

**SECTION-D**  
**2Qx15M= 30 Marks**

- 1. Each Question carries 15 Marks.**  
**2. Instruction: Write a long answer.**

Q1.	<p>The following Table below shows the transactions for a grocery store. Find out frequent item sets considering support count threshold = 4 and interesting association rules at 80% confidence level. [Item sold for a particular transaction is shown with a ✓ mark].</p> <table border="1" style="width: 100%; text-align: center;"> <thead> <tr> <th>Transactionl D</th> <th>Eggs</th> <th>Salt</th> <th>Butter</th> <th>Bread</th> <th>Milk</th> <th>Rice</th> </tr> </thead> <tbody> <tr><td>1.</td><td>✓</td><td>✓</td><td>✓</td><td>✓</td><td></td><td></td></tr> <tr><td>2.</td><td></td><td></td><td>✓</td><td>✓</td><td>✓</td><td></td></tr> <tr><td>3.</td><td></td><td></td><td></td><td>✓</td><td></td><td>✓</td></tr> <tr><td>4.</td><td>✓</td><td></td><td>✓</td><td></td><td></td><td></td></tr> <tr><td>5.</td><td></td><td></td><td>✓</td><td>✓</td><td>✓</td><td></td></tr> <tr><td>6.</td><td>✓</td><td>✓</td><td>✓</td><td>✓</td><td></td><td></td></tr> <tr><td>7.</td><td></td><td></td><td></td><td>✓</td><td></td><td>✓</td></tr> <tr><td>8.</td><td></td><td></td><td>✓</td><td>✓</td><td></td><td></td></tr> <tr><td>9.</td><td>✓</td><td></td><td>✓</td><td>✓</td><td>✓</td><td></td></tr> <tr><td>10.</td><td>✓</td><td>✓</td><td>✓</td><td>✓</td><td></td><td></td></tr> </tbody> </table>	Transactionl D	Eggs	Salt	Butter	Bread	Milk	Rice	1.	✓	✓	✓	✓			2.			✓	✓	✓		3.				✓		✓	4.	✓		✓				5.			✓	✓	✓		6.	✓	✓	✓	✓			7.				✓		✓	8.			✓	✓			9.	✓		✓	✓	✓		10.	✓	✓	✓	✓			<b>CO3</b>
Transactionl D	Eggs	Salt	Butter	Bread	Milk	Rice																																																																									
1.	✓	✓	✓	✓																																																																											
2.			✓	✓	✓																																																																										
3.				✓		✓																																																																									
4.	✓		✓																																																																												
5.			✓	✓	✓																																																																										
6.	✓	✓	✓	✓																																																																											
7.				✓		✓																																																																									
8.			✓	✓																																																																											
9.	✓		✓	✓	✓																																																																										
10.	✓	✓	✓	✓																																																																											
Q2.	The following Table-below presents a training set, D, of class-labeled tuples randomly selected from the customer database. IF X= (Age = Youth, Income = Medium, Student =	<b>CO3</b>																																																																													

Yes, Credit-rating=excellent). Calculate the Information-Gain & Entropy for all attributes for splitting criterion for the given training-set Table below.

RID	Age	Income	Student	Credit-rating	Class: BuyComputer
1	Youth	High	No	Fair	No
2	Youth	High	No	Excellent	No
3	Middle-Aged	High	No	Fair	Yes
4	Senior	Medium	No	Fair	Yes
5	Senior	Low	Yes	Fair	Yes
6	Senior	Low	Yes	Excellent	No
7	Middle-Aged	Low	Yes	Excellent	Yes
8	Youth	Medium	No	Fair	No
9	Youth	Low	Yes	Fair	Yes
10	Senior	Medium	Yes	Fair	Yes
11	Youth	Medium	Yes	Excellent	Yes
12	Middle-Aged	Medium	No	Excellent	Yes
13	Middle-Aged	High	Yes	Fair	Yes
14	Senior	Medium	No	Excellent	No