Parshvanath Charitable Trust's
A. P. SHAH INSTITUTE OF TECHNOLOGY
(Approved by AICTE New Delhi & Govt. of Maharashtra, Affiliated to University of Mumbai)
(Religious Jain Minority)

DEPARTMENT OF INFORMATION TECHNOLOGY

UNIT TEST-I

Class: TE **Semester: VI** **Subject: DMBI**

Date: **Time:** **Max marks: 40**

Note the following instructions

1. **Attempt all questions.**
2. **Draw neat diagrams wherever necessary.**
3. **Write everything in Black ink (no pencil) only.**
4. **Assume data, if missing, with justification.**

Q.N	Questions	MARKS	CO	Blooms Taxonomy Level	PO2
Q.1.	**Attempt any two.**				
	a) Explain various features of Data Warehouse?	[5]	CO1	L2	
	b) Explain Star and Snowflakes schemas with example.	[5]	CO1	L2	
	c) Describe KDD process using figure	[5]	CO1	L2	
	d) Compare in between OLTP and OLAP	[5]	CO1	L2	
Q.2.	**Attempt any two**				
	a) Solve the Partition the given data into 4 bins using Equi-depth binning method and perform smoothing according to the following methods. 1. Smoothing by bin mean 2. Smoothing by bin median 3. Smoothing by bin boundaries Data: 4, 8, 9, 15, 21, 21, 24, 25, 26, 28, 29, 34	[10]	CO2	L3	PO2, PO12

<table>
<tr><td></td><td>b) Solve and fine Mean, Median and mode for a given data and show plot.
Data:
11,13,13,15,15,16,19,20,20,21,21,22,23,24,30,40,45,45,45</td><td>[10]</td><td>CO2</td><td>L3</td><td>PO2, PO12</td></tr>
<tr><td></td><td>c) Solve, suppose that data for analysis includes the attribute age. The age values for data tuples are (in increasing order):
13,15,16,16,19,20,20,21,22,22,25,25,25,25,30,33,33,35,35,35,35,36,40,45,46,52,70
1. What is mean of data? What is median of data?
2. What is mode of data? Comment on data's modality.
3. What is mid-range of data?</td><td>[10]</td><td>CO2</td><td>L3</td><td>PO2, PO12</td></tr>
<tr><td>Q.3.</td><td>Attempt any one.</td><td></td><td></td><td></td><td></td></tr>
<tr><td></td><td>a) Solve the problem using the given training dataset classify the following tuple using Naïve Bayes Algorithm:
<table><tr><th>Home Owner</th><th>Marital Status</th><th>Annual Income</th><th>Defaulted Borrower</th></tr><tr><td>Yes</td><td>Single</td><td>125K</td><td>No</td></tr><tr><td>No</td><td>Married</td><td>100K</td><td>No</td></tr><tr><td>No</td><td>Single</td><td>70K</td><td>No</td></tr><tr><td>Yes</td><td>Married</td><td>120K</td><td>No</td></tr><tr><td>No</td><td>Divorced</td><td>95K</td><td>Yes</td></tr><tr><td>No</td><td>Married</td><td>60K</td><td>No</td></tr><tr><td>Yes</td><td>Divorced</td><td>220K</td><td>No</td></tr><tr><td>No</td><td>Single</td><td>85K</td><td>Yes</td></tr><tr><td>No</td><td>Married</td><td>75K</td><td>No</td></tr><tr><td>No</td><td>Single</td><td>90K</td><td>Yes</td></tr></table>
Given the test record X = (Home Owner = No, Marital status=Married, Annual Income =120K)
Compute its class label using Naïve Bayes Classifier</td><td>[10]</td><td>CO3</td><td>L3</td><td>PO2,
PO5,PO12</td></tr>
<tr><td></td><td>b) Apply confusion matrix to calculate Accuracy, Precision and Recall for the following Confusion data
<table><tr><th>Cancer Classes</th><th>Yes</th><th>No</th><th>Total</th></tr><tr><td>Yes</td><td>90</td><td>210</td><td>300</td></tr><tr><td>No</td><td>140</td><td>9560</td><td>9700</td></tr><tr><td>Total</td><td>230</td><td>9770</td><td>10000</td></tr></table></td><td>[10]</td><td>CO3</td><td>L3</td><td>PO2,
PO5,PO12</td></tr>
</table>