SET I

Name: Enrolment No:	**UPES** UNIVERSITY WITH A PURPOSE

<table>
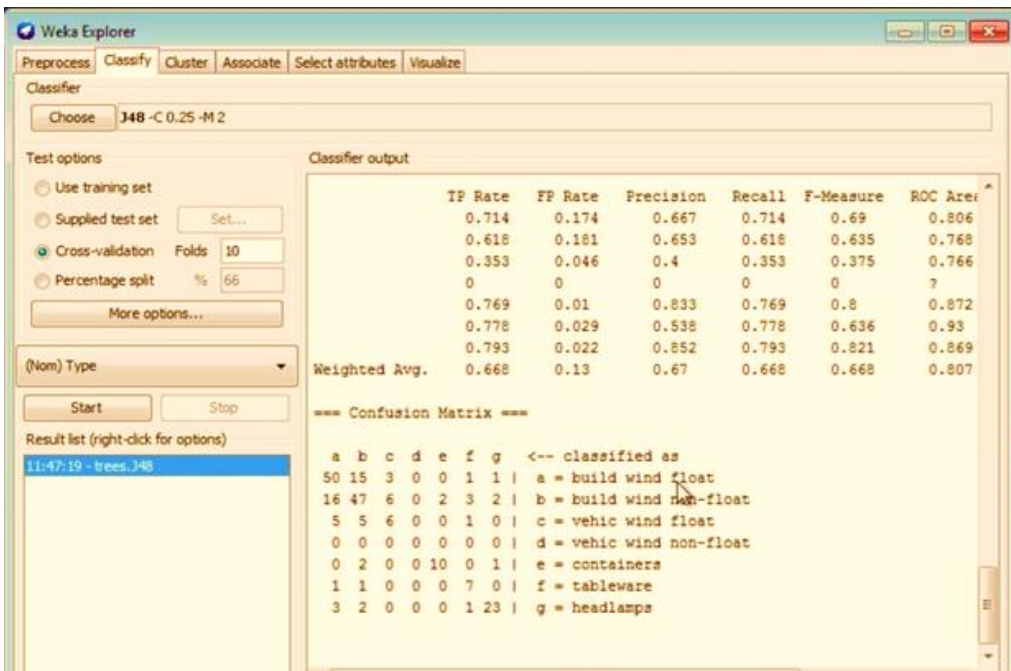<tr><td colspan="4">UNIVERSITY OF PETROLEUM AND ENERGY STUDIES
End Semester Examination, May 2019</td></tr>
<tr><td colspan="2">Course: Data Mining
Program: MBA(BA)
Course code: DSBA 7008</td><td colspan="2">Semester: II
Time: 03 Hours
Max. Marks: 100</td></tr>
<tr><td colspan="4">Instruction: Use Weka to solve questions wherever required.</td></tr>
<tr><td colspan="4" align="center">SECTION A</td></tr>
<tr><td></td><td></td><td>Marks</td><td>CO</td></tr>
<tr><td></td><td>Use Weka to answer the below questions:</td><td></td><td></td></tr>
<tr><td>Q1</td><td>Open the <u>contact-lenses</u> dataset.

a) How many instances are there?
b) How many attributes are there?
c) How many possible values are there for the <i>age</i> attribute?
d) Which of these attributes has <i>reduced</i> as a possible value?
e) What is the default name of class attribute?</td><td>5</td><td>CO2</td></tr>
<tr><td></td><td></td><td></td><td></td></tr>
<tr><td colspan="4" align="center">SECTION B</td></tr>
<tr><td>Q 2.</td><td>Differentiate between the following:

a) Classification and association
b) Classification and regression
c) Prune and unprune</td><td>(3X5=
15)</td><td>CO2</td></tr>
<tr><td>Q3.</td><td>Define following with example:

a) Cross validation
b) Percentage split
c) Training data
d) Clustering
e) Test data</td><td>(5X5=
25)</td><td>CO1</td></tr>
<tr><td>Q4.</td><td>Describe the below given confusion matrix:</td><td>5</td><td>CO2</td></tr>
</table>

Weighted Avg.	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area
	0.714	0.174	0.667	0.714	0.69	0.806
	0.618	0.181	0.653	0.618	0.635	0.768
	0.353	0.046	0.4	0.353	0.375	0.766
	0	0	0	0	0	?
	0.769	0.01	0.833	0.769	0.8	0.872
	0.778	0.029	0.538	0.778	0.636	0.93
	0.793	0.022	0.852	0.793	0.821	0.869
Weighted Avg.	0.668	0.13	0.67	0.668	0.668	0.807

```
=== Confusion Matrix ===

  a  b  c  d  e  f  g   <-- classified as
 50 15  3  0  0  1  1 |  a = build wind float
 16 47  6  0  2  3  2 |  b = build wind non-float
  5  5  6  0  0  1  0 |  c = vehic wind float
  0  0  0  0  0  0  0 |  d = vehic wind non-float
  0  2  0  0 10  0  1 |  e = containers
  1  1  0  0  0  7  0 |  f = tableware
  3  2  0  0  0  1 23 |  g = headlamps
```

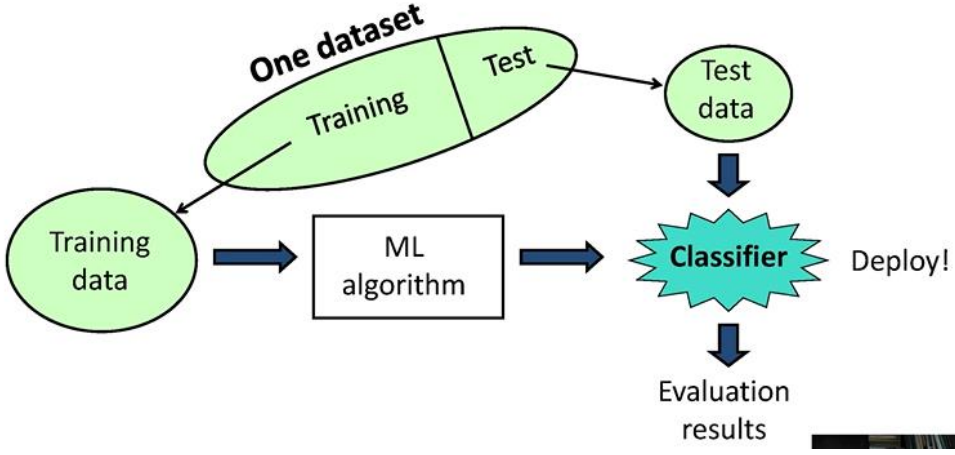|			
	SECTION-C		

Q5.	**Classify the attribute 'Type' of glass.arff dataset using J48.**
	    a) No. of instances and Attributes
	    b) No of leaves and trees
	    c) Overall accuracy
	    d) Write confusion matrix
	    e) Write impact when Change minNumObj=15
	    f) Display Decision tree
	    g) Write interpretation of Model
		(7X2= 14)	**CO3**

Q6.	**Open weather.normal.arff file.**
	    a) Write step to remove 3rd attribute
	    b) Write step to remove High values of Humidity attribute
		(2X3= 6)	**CO3**

	SECTION-D		

Q7.	**Open cpu.arff file.**
	    a) Run linear regression.
	    b) Write linear equation.
		(2X5= 10)	**CO3**

	c) Run Non- linear regression. d) Write all the linear equations. e) Justify which one is better and why.		
Q8.	Define K mean clustering algorithm and run the weather.numeric.arff data for clustering and write the output with justification.	**(10)**	**CO2**
Q9.	Run Apriori algorithm on weather.nominal data and write the ten best association rules.	**(10)**	**CO3**

SET II

UNIVERSITY OF PETROLEUM AND ENERGY STUDIES **End Semester Examination, May 2019**	
Course: Data Mining **Program: MBA(BA)** **Course code: DSBA 7008**	**Semester: II** **Time: 03 Hours** **Max. Marks: 100**

Instructions: Use Weka to solve questions wherever required.

		Marks	CO
	SECTION A		

Q1	**Write short notes on the following:**
	  a) Weka
	  b) Attribute
	  c) Instances
	  d) Flat file
	  e) J48
		5	**CO1**

	SECTION B		

Q 2.	**Differentiate between the following:**
	
	  a) Classification and association
	  b) Nominal and Numeric attribute
	  c) Training and Test data
		(3X5= 15)	**CO2**

Q3.	**Define following with example:**
	
	  a) Classification
	  b) Association
	  c) Regression
	  d) Clustering
	  e) Overfitting
		(5X5= 25)	**CO1**

Q4.	**Explain the below given diagram:** 	**(5)**	**CO2**
	SECTION-C		
Q 5.	**Classify the attribute 'Type' of glass.arff dataset using J48.** a) No. of instances and Attributes b) No of leaves and trees c) Overall accuracy d) Write confusion matrix e) Write impact when Change minNumObj=15 f) Display Decision tree g) Write interpretation of Model	**(7X2= 14)**	**CO2**
Q6.	**Open weather.normal.arff file.** a) Write step to remove 3rd attribute b) Write step to remove High values of Humidity attribute	**(2X3= 6)**	**CO2**
	SECTION-D		
Q7.	**Open cpu.arff file.** a) Run linear regression. b) Write linear equation.	**(2X5= 10)**	**CO3**

<table>
<tr><td></td><td>c) Run Non- linear regression.
d) Write all the linear equations.
e) Justify which one is better and why.</td><td></td><td></td></tr>
<tr><td>Q8.</td><td>**Answer the following based on below given file:**

@relation weather.symbolic

@attribute outlook {sunny, overcast, rainy}
@attribute temperature {hot, mild, cool}
@attribute humidity {high, normal}
@attribute windy {TRUE, FALSE}
@attribute play {yes, no}

@data
sunny,hot,high,FALSE,no
sunny,hot,high,TRUE,no
overcast,hot,high,FALSE,yes
rainy,mild,high,FALSE,yes
rainy,cool,normal,FALSE,yes
rainy,cool,normal,TRUE,no
overcast,cool,normal,TRUE,yes
sunny,mild,high,FALSE,no
sunny,cool,normal,FALSE,yes
rainy,mild,normal,FALSE,yes
sunny,mild,normal,TRUE,yes
overcast,mild,high,TRUE,yes
overcast,hot,normal,FALSE,yes
rainy,mild,high,TRUE,no

a) Write the name of relation, attributes and number of instances.
b) Name the class attribute and its labels.
c) Write any two decisions list based on above data.
d) Calculate number of possible instances.
e) Draw decision tree based on above data.</td><td>**CO2**

3
2
5
3
7</td><td></td></tr>
</table>