Government Polytechnic Pune – 411016

(An Autonomous Institute Of Government of Maharashtra)

Progressive Assessment Test No- 1

Programme : Computer
Diploma : Regular
Course Code : CM5107
Duration : 1hour (12.30pm to 1.30pm)
No of Students Registered for this Course : 219

Term : EVEN 2021
Course Name : Data Mining
Maximum Marks : 20
Date : 28/04/2022

Instructions: 1) Figures to the right indicate full marks. 2) Assume suitable data if necessary. 3)Draw neat sketches wherever applicable. 4) Follow all the examination rules.

CO	LEVEL R/U/A		Q. No. 1 Attempt any Four	Marks 2*4 =08
CM5107.1	R	a)	Define Data Mining . List any two applications of Data Mining.	
CM5107.2	U	b)	Differentiate between discrete and continuous attributes.	
CM5107.3	R	c)	State the tasks used for Data Preprocessing.	
CM5107.1	R	d)	Enlist different kinds of data. State the use of any two types.	
CM5107.2	R	e)	Define following terms : (a) Mean (b) Median (c) Mode (d) IQR	
CM5107.3	R,U	f)	List down the methods to handle missing values in Data Cleaning?	

CO	LEVEL R/U/A		Q. No. 2 Attempt any Three	Marks 4*3 =12
CM5107.1	R	a)	Write a short note on cluster analysis with suitable example.	
CM5107.2	A	b)	You grow 7 crystals from a solution and measure the length of each crystal in millimeters. Here is your data:600,470,170,430,300,290,520. Calculate the sample standard deviation of the length of the crystals.	
CM5107.3	A	c)	Remove noisy data from following data set by using Binning methods using all smoothing techniques. 2, 6, 7, 9, 13, 20, 21, 24, 30	
CM5107.1	U	d)	Describe the following data mining patterns: (a) Outlier Analysis (b) Mining frequent patterns.	
CM5107.2	U	e)	Explain the following terms with suitable examples: (a) Binary attribute (b) Numeric attribute.	

Government Polytechnic Pune – 411016

(An Autonomous Institute Of Government of Maharashtra)

Progressive Assessment Test No- 2

Programme : Computer
Diploma : Regular
Course Code : CM5107
Duration : 1hour
No of Students Registered for this Course : 219

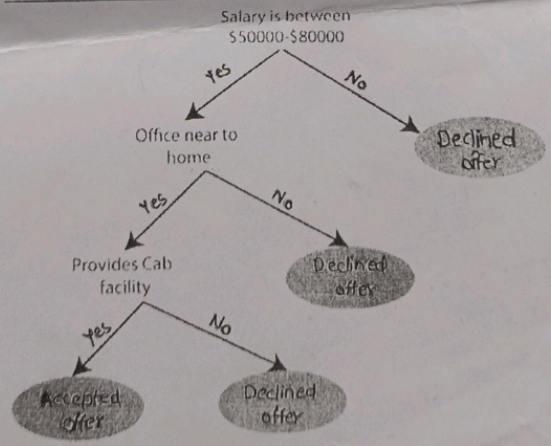
Term : EVEN 2021
Course Name : Data Mining
Maximum Marks : 20
Date : /06/2022

Instructions: 1) Figures to the right indicate full marks. 2) Assume suitable data if necessary. 3)Draw neat sketches wherever applicable. 4) Follow all the examination rules.

CO	LEVEL R/U/A		Q. No. 1 Attempt any Four	Marks 2*4 =08
CM5107.4	R,U	a)	Define classification in Data Mining with suitable examples.	2
CM5107.5	R	b)	List any 4 open source data mining tools.	2
CM5107.6	U	c)	State the difference between data mining and data warehouse.	2
CM5107.4	R	d)	Define the following terms:- (1) Maximum margin (2) Positive Hyperlane (3) Negative Hyperlane (4) MMH	2
CM5107.5	U	e)	Write a short note on ARFF.	2
CM5107.6	R,U	f)	Enlist any four types of classification techniques.	2

CO	LEVEL R/U/A		Q. No. 2 Attempt any Three	Marks 4*3 =12
CM5107.6	U	a)	Differentiate between Operational Database Systems and Data Warehouses	4
CM5107.4	A	b)	Draw decision tree from the given table :-	4

Outlook	Temperature	Humidity	Windy	Play Golf
Rainy	Hot	High	False	No
Rainy	Hot	High	True	No
Overcast	Hot	High	False	Yes
Sunny	Mild	High	False	Yes
Sunny	Cool	Normal	False	Yes
Sunny	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Rainy	Mild	High	False	No
Rainy	Cool	Normal	False	Yes

CM5107.4	A	c)	Write an algorithm for Classification by back propagation.	4
CM5107.4	A	d)	Form classification rule from the following decision tree : 	4
CM5107.5	U	e)	Enlist features of Weka Tool.	4

Enroll. No.

GOVERNMENT POLYTECHNIC, PUNE

(AN AUTONOMOUS INSTITUTE OF GOVT. OF MAHARASHTRA)

TERM EXAM – EVEN 2021

Programme: Diploma in Computer Engineering
Course Name: Data Mining
Course Code: CM5107

Term: Even 2021
Time Allowed: 3 hrs.
Marks: 80

INSTRUCTIONS :

1. All questions are compulsory.
2. Illustrate your answer with neat sketches wherever necessary.
3. Use of block pattern, slide rule, Mathematical and Steam tables, Non-programmable electronic pocket calculator and Moilers chart is not permissible.
4. Figure to the right indicate full marks.
5. Assume suitable additional data if necessary.
6. Possession of mobile phone is strictly prohibited during the examinations.
7. Answer each next main Question on a new page.
8. The CO & Level (R, U, A) mentioned in the extreme right columns are for academic purpose only.

SECTION -I

	Q. No. 1 Attempt any FOUR of the following **(4*4=16)**	**CO**	**Level (R,U,A)**
a)	Define Data Mining and state any four uses of it in the real world.	CO1	R,U
b)	Find Standard Deviation for the following data sets : 9,2,5,4,12,7,8,11,9,3,7,4,12,5,4,10,9,6,9,4	CO2	A
c)	Explain Attribute Subset Selection with suitable examples.	CO3	U
d)	Write a short note on Outlier Analysis with example.	CO1	R
e)	Describe IQR and Quartiles in dispersion with suitable example.	CO2	U
f)	Write the steps used for Data Integration and Transformation and explain them.	CO3	U

	Q. No. 2 Attempt any THREE of the following **(4*3=12)**		
a)	Differentiate between Data Warehouse and Data Mining.	CO1	R,U
b)	Define following attribute with example: (1) Binary Attribute. (2) Nominal Attribute	CO2	R
c)	List Various types of Data Reduction. Explain any two ways of Data Reduction with suitable examples.	CO3	U
d)	Give example for the following databases: (1) Multimedia (2) Heterogeneous (3) World Wide Web (4) Data Streams	CO1	A
e)	Write a short note on Principal Component Analysis with suitable examples.	CO3	R,U

	Q. No. 3 Attempt any THREE of the following **(4*3=12)**		
a)	Explain Cluster Analysis with suitable examples.	CO1	U
b)	Differentiate between Qualitative and Quantitative Attributes.	CO2	R,U
c)	Write a short note on Wavetlet Transforms with suitable examples.	CO3	U
d)	There are different measures of Dispersions used in Data Mining, which measure of dispersion is the best? Explain with reasons and examples.	CO2	U,A
e)	Explain Histogram in Data Reduction with suitable examples.	CO3	U

P.T.O.→

Section - II

			CO	Level (R,U,A)
Q. No. 4 Attempt any FOUR of the following		**(4*4= 16)**		
a)	Describe need of Genetic algorithm in data mining.		CO4	R
b)	Write a short note on following:			
	i)	Classification by back propagation	CO4	R
	ii)	Rule based classification		
c)	Describe KNN- classifier with example.		CO4	U
d)	State use of Data Warehouse. Describe ETL in Data Warehouse.		CO6	R,U
e)	Differentiate between Database and Data Warehouse.		CO6	R
f)	Elaborate the stages of Data warehouse with suitable diagram.		CO6	R
Q. No. 5 Attempt any THREE of the following		**(4*3= 12)**		
a)	Describe Cluster Analysis with neat diagram.		CO4	U
b)	State applications of Fuzzy Set Theory.		CO4	A
c)	Explain different functions of WEKA TOOL.		CO5	A
d)	Write an algorithm for Decision tree.		CO4	A
e)	Describe the need of Classification.		CO6	U
Q. No. 6 Attempt any THREE of the following		**(4*3= 12)**		
a)	Elaborate Support vector machines with diagram.		CO4	A
b)	State any four advantages of K-means algorithm.		CO4	A
c)	Describes any two machine learning algorithm used in WEKA TOOL.		CO5	R
d)	Differentiate between OLTP and OLAP.		CO6	A
e)	Explain multidimensional Data model? How it is used in data warehouse?		CO6	U