Paper / Subject Code: BE1356 / OPEN ELECTIVE Data Analytics

BE1356

Total No. of Printed Pages: 3

B.E. (Information Technology) Semester-VII (Revised Course 2019-20) EXAMINATION JANUARY 2023

Open Elective: - Data Analytics

[Time: Three Hours]

[Max. Marks: 199]

Instructions:

1) Figures to the right indicate full Marks

2. Answer any Five questions by selecting Two questions from Part A and Two from Part B

and each One question from Part C

Part -A

Answer any Two questions from the following:

Q1	 a) Explain 3 attributes that define big data characteristics. b) What are the reasons for current data deluge (flooding)? c) What are the business drivers for advanced analytics? 		6 8 6
Q2	Explain supervised learning and unsupervised learning. Using following data construct a decision tree for label = "buy_computer"		6 8

Age	Income	Student	credit_rating	buy_computer
(-30)	(righ	no <	fair /	no /
(-31)	(II)	no /	excellent &	no /
3140	(19)	00 /	fair /	yes /
>40 V	mid 4	no /	fair /	yes /
>40 M	low /	yes	fair /	yes /
>40 V	low /	yes	excellent &	no /
31400	low /	yes	excellent &	yes f.
(3)	mid ⊀	no /	fair /	no - p
=30	low /	yes	fair /	yes /
40 V	mid *	yes	fair -	yes /
39	mid 🛠	yes	excellent &	yes /
1400	mid 🐰	no /	excellent (yes «
140	(igh	ýes	fair /	yes /
40 V	medien *	no /	excellent &	100

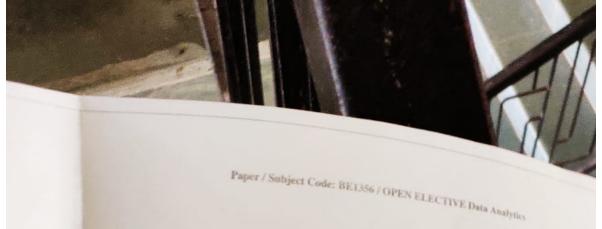
Paper / Subject Code: BE1356 / OPEN ELECTIVE Data Analytics

BE1356

	6) Explain any hierarchical method of clustering .	6
(3)	Explain key roles for a successful analytics project. Identify and explain main phases of Data Analytics Lifecycle. Briefly explain R graphical user interfaces.	6 8 6
	Part -B	
A	nswer any Two questions from the following:	
Q4	Why Frequent Pattern Mining is important? With a simple example explain the Apriori algorithm for frequent pattern mining. Explain any 3 applications of Anomaly Detection.	6 8 6
QS	 (a) Explain 3 important steps of text analysis problem. (b) Briefly explain Term Frequency - Inverse Document Frequency (TFIDF) measure. (c) Explain MAP Reduce paradigm 	6 8 6
Q6	a) Explain Hadoop Distributed File System. b) List and briefly explain any 2 Hadoop-related Apache projects. c) Write a brief note on NoSQL.	6 8 6
	Part -C	
	Answer any One question from the following:	

pal length in cm		al length in cm	l width in	class:
4.9	3	1.4	0,2	Iris-setosa
4.6	3.1	1.5	0.2	Iris-setosa
5.4	3.9	1.7	0.4	Iris-setosa
5	3.4	1.5	0.2	Iris-setosa
4.9	3.1	1.5	0.1	Iris-setosa
6.9	3.1	4.9	1.5	Iris-versicolor
6.5	2.8	4.6	1.5	Iris-versicolor
6.3	3.3	4.7	1.6	Iris-versicolor

a) Using K-nearest neighbour algorithm and below data classify sample X (6.5, 3.0, 2.4, 1.9). Take K 8



BE1356

Distance	Tag.	men Ke	gression	Equation Equation	in for vi-			
b) Determ Distance Price	10	12	14	15	17	e tollow	ing data.	
Price	800	700	600	+		10	25	30
		1	1000	550	500	450	400	200

- A) Give some examples of Big Duta Analytics.
- (28 a) Explain by 3 use cases of linear regression.
 b) Explain model description for Logistic Regression.
 c) Why Frequent Pattern Mining is important?

Explain appl=g)