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:100]**

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. 6
b) What are the reasons for current data deluge (flooding)? 8
c) What are the business drivers for advanced analytics? 6

Q2 a) Explain supervised learning and unsupervised learning. 6
b) Using following data construct a decision tree for label = "buy_computer" 8

Age	Income	Student	credit_rating	buy_computer
<=30	high	no	fair	no
<=30	high	no	excellent	no
31…40	high	no	fair	yes
>40	mid	no	fair	yes
>40	low	yes	fair	yes
>40	low	yes	excellent	no
31…40	low	yes	excellent	yes
<=30	mid	no	fair	no
<=30	low	yes	fair	yes
>40	mid	yes	fair	yes
<=30	mid	yes	excellent	yes
31…40	mid	no	excellent	yes
31…40	high	yes	fair	yes
>40	medium	no	excellent	no

c) Explain any hierarchical method of clustering . 6

Q3 a) Explain key roles for a successful analytics project. 6
b) Identify and explain main phases of Data Analytics Lifecycle. 8
c) Briefly explain R graphical user interfaces. 6

Part –B

Answer any Two questions from the following:

Q4 a) Why Frequent Pattern Mining is important? 6
b) With a simple example explain the Apriori algorithm for frequent pattern mining. 8
c) Explain any 3 applications of Anomaly Detection. 6

Q5 a) Explain 3 important steps of text analysis problem. 6
b) Briefly explain Term Frequency - Inverse Document Frequency (TFIDF) measure. 8
c) Explain MAP Reduce paradigm 6

Q6 a) Explain Hadoop Distributed File System. 6
b) List and briefly explain any 2 Hadoop-related Apache projects. 8
c) Write a brief note on NoSQL. 6

Part –C

Answer any One question from the following:

Q7 a) Using K-nearest neighbour algorithm and below data classify sample X (6.5, 3.0, 2.4, 1.9). Take K = 5. 8

se pal length in cm	s epal width in cm	pet al length in cm	peta l width in cm	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
6.6	2.9	4.6	1.3	Iris-versicolor
5	2	3.5	1	Iris-versicolor
6.5	3	5.8	2.2	Iris-virginica
4.9	2.5	4.5	1.7	Iris-virginica
7.2	3.6	6.1	2.5	Iris-virginica
6.8	3	5.5	2.1	Iris-virginica
5.8	2.8	5.1	2.4	Iris-virginica

b) Determining Linear Regression Equation for the following data.

Distance	10	12	14	15	17	18	25	30
Price	800	700	600	550	500	450	400	200

c) Give some examples of Big Data Analytics.

Q8 a) Explain ay 3 use cases of linear regression.
b) Explain model description for Logistic Regression.
c) Why Frequent Pattern Mining is important?

Explain appl = g