**K. J. Somaiya Institute of Technology, Sion, Mumbai-22**
**(Autonomous College Affiliated to University of Mumbai)**

May-June 2024
Program: _ B.Tech_ Scheme I/II/IIB/III:_II_
Regular/Supplementary Examination: LY Semester: VIII
Course Code: **UEXDLC8034** and Course Name: Fundamental of Data Science

Date of Exam: 18/05/2024          Duration: 02.5  Hours          Max. Marks: 60

Instructions:
(1) All questions are compulsory.
(2) Draw neat diagrams wherever applicable.
(3) Assume suitable data, if necessary.

| | | Max. Marks | CO | BT level |
|---|---|---|---|---|
| **Q 1** | **Solve any six questions out of eight:** | **12** | | |
| i) | Explain the Pivot operation in brief. | 2 | 1 | U |
| ii) | Define outlier | 2 | 2 | U |
| iii) | Explain the term Data Validation | 2 | 3 | U |
| iv) | Explain the Roll up operation | 2 | 4 | U |
| v) | A fact Table is deep Explain this statement | 2 | 4 | U |
| vi) | List various schemes for warehouse designing | 2 | 5 | U |
| vii) | Write Short note on Extended version of ARIMA models | 2 | 6 | U |
| viii) | List & Explain in a line, about the types of Clustering Methods? | 2 | 4,5 | U |
| **Q.2** | **Solve any four questions out of six.** | **16** | | |
| i) | Explain how to sort a Data Frame with respect to multiple columns? | 4 | 1 | U |
| ii) | Discuss how the exceptions defined. Explain how it is handled with example | 4 | 2 | U |
| iii) | Explain Fuzzy set Approach for Classification | 4 | 3 | U |
| iv) | Explain Frequent Pattern Mining | 4 | 4 | U |
| v) | State Issues Regarding Classification and Prediction | 4 | 5 | U |
| vi) | Write Short note on Autoregressive with exogenous(ARX) | 4 | 6 | U |

| | |
|---|---|
| May-June 2024 | |
| Program: _ B.Tech_ Scheme I/II/IIB/III:_II_ | |
| Regular/Supplementary Examination: LY Semester: VIII | |
| Course Code: ▆EXDLC8034 and Course Name: Fundamental of Data Science | |
| Date of Exam: 18/05/2024 | Duration: 02.5 Hours | Max. Marks: 60 |

**Q.3**            **Solve any two questions out of three.**        **16**

i)    Write syntaxes for instructions used to read & write Date in text format with    8    1    U
example.

ii)    Use the k-means algorithm and Euclidean distance to cluster the following 8    8    3    A
examples into 3 clusters:
A1=(2,10), A2=(2,5), A3=(8,4), A4=(5,8), A5=(7,5), A6=(6,4), A7=(1,2), A8=(4,9).
The distance matrix based on the Euclidean distance is given below:

| | A1 | A2 | A3 | A4 | A5 | A6 | A7 | A8 |
|---|---|---|---|---|---|---|---|---|
| A1 | 0 | $\sqrt{25}$ | $\sqrt{36}$ | $\sqrt{13}$ | $\sqrt{50}$ | $\sqrt{52}$ | $\sqrt{65}$ | $\sqrt{5}$ |
| A2 | | 0 | $\sqrt{37}$ | $\sqrt{18}$ | $\sqrt{25}$ | $\sqrt{17}$ | $\sqrt{10}$ | $\sqrt{20}$ |
| A3 | | | 0 | $\sqrt{25}$ | $\sqrt{2}$ | $\sqrt{2}$ | $\sqrt{53}$ | $\sqrt{41}$ |
| A4 | | | | 0 | $\sqrt{13}$ | $\sqrt{17}$ | $\sqrt{52}$ | $\sqrt{2}$ |
| A5 | | | | | 0 | $\sqrt{2}$ | $\sqrt{45}$ | $\sqrt{25}$ |
| A6 | | | | | | 0 | $\sqrt{29}$ | $\sqrt{29}$ |
| A7 | | | | | | | 0 | $\sqrt{58}$ |
| A8 | | | | | | | | 0 |

Suppose that the initial seeds (centers of each cluster) are A1, A4 and A7. Run the
k-means algorithm for
1 epoch only. At the end of this epoch show:
a) The new clusters (i.e. the examples belonging to each cluster)
b) The centers of the new clusters
c) Draw a 10 by 10 space with all the 8 points and show the clusters after the first
epoch and the new
centroids.
d) How many more iterations are needed to converge? Draw the result for each
epoch.

iii)    Write and explain about Classification by Back propagation Algorithm.    8    5    U

**K. J. Somaiya Institute of Technology, Sion, Mumbai-22**
**(Autonomous College Affiliated to University of Mumbai)**

| | |
|---|---|
| May-June 2024 | |
| Program: _ B.Tech_ Scheme I/II/IIB/III:_II_ | |
| Regular/Supplementary Examination: LY Semester: VIII | |
| Course Code: EXDLC8034 and Course Name: Fundamental of Data Science | |
| Date of Exam: 18/05/2024     Duration: 02.5 Hours     Max. Marks: 60 | |

**Q.4**          **Solve any two questions out of three.**      **16**

| | | | | |
|---|---|---|---|---|
| i) | Suppose that the data for analysis includes the attribute salary. We have the following values of salary (in Thousands of INR), shown in increasing order: 30, 36, 47, 50, 52, 52, 56, 60, 63, 70, 70, 110. <br> a) What are the mean, median, mode and midrange of the data? <br> b) Find first quartile, second quartile and third quartile of the data. <br> c) Show the boxplot of the data | 8 | 2 | U |
| ii) | Write the syntax and explain in a line about these operations in pandas: <br><br> i.     Sorting, Group by operation using Pandas <br> ii     Give the syntax and explain in a line, 4 essentials basic methods that are used to get some information about the data frame | 8 | 1,2,3 | A |
| iii) | Explain Autoregressive Moving Average(ARMAX) along with diagrammatic representation. | 8 | 6 | U |

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*