



M.Sc. III Semester Degree Examination, April/May - 2024

COMPUTER SCIENCE

Data Analytics

(NEP)

Time : 3 Hours

Maximum Marks : 70

Note : Answer **any five** of the following questions with Question **No.1** is **Compulsory**, each question carries **equal** marks.

1. (a) Differentiate between Data, Big Data, and Data Science. How do they interrelate in modern data-driven environments ? **7**
(b) Explain KDD process in detail. **7**
2. (a) Explain the concept of infographics and word clouds as tools for visualizing multivariate data. Provide examples of situations where infographics and word clouds are effective in conveying complex information and trends. **7**
(b) Describe the process of converting data to a different scale in statistical analysis. Give an example. **7**
3. (a) Apply the DBSCAN algorithm for the following data points and create clusters with a min points 4 and epsilon 1.9. **10**
P1(3, 7), P2(4, 6), P3(5, 5), P4(6, 4), P5(7, 3), P6(6, 2), P7(7, 2), P8(8,4), P9(3, 3), P10(2, 6), P11(3, 5), P12(2, 4).
(b) Write an algorithm for K-means clustering technique. **4**
4. (a) Briefly explain with an example why should we use different data to train and to test a model ? **7**
(b) Discuss the importance of descriptive analytics in deriving insights and making data driven decisions. **7**
5. (a) Define binary classification and provide examples of real-world applications where binary classification is commonly used. **7**
(b) Explain Naive-bayes algorithm. **7**
6. (a) Construct Box Plot diagram for the following Age attribute. **7**
70, 75, 86, 73, 89, 84, 82, 56, 68, 69, 80, 94, 64, 58, 53, 68, 99.
(b) Compare and contrast two clustering techniques suitable for analyzing the healthcare organization's patient data. Discuss the strengths and limitations of each technique and recommend the most appropriate approach for this specific healthcare scenario. **7**



7. (a) Explain linear regression technique. Mention its advantages and dis-advantages. **7**
- (b) Write the pseudo-code for the k-NN algorithm. List out its advantages and dis-advantages. **7**
8. Write short notes on the following : **5+5+4**
- (a) CRISP-DM Methodology
- (b) Apriori algorithm
- (c) Handling missing values

- o o o -

