

(4) . MB-203(BA3)

of groups so that the data points within each group are more comparable to one another and different from the data points within the other groups. It is essentially a grouping of things based on how similar and different they are to one another.

We are given a data set of items, with certain features, and values for these features (like a vector). The task is to categorize those items into groups. To achieve this, we will use the K means algorithm; -an unsupervised learning algorithm. 'K' in the name of the algorithm represents the number of groups/clusters we want to classify our items into.

Q : Based on K-means cluster the following eight points with (x, y) representing locations into three 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)

Write is the meaning of unlabeled data in the above-mentioned case.

MB-203(BA3)

480

H

Roll No.

MB-203(BA3)

**M. B. A. (SECOND SEMESTER)
END SEMESTER EXAMINATION,**

June, 2023

DATA MINING

Time : Three Hours

Maximum Marks : 100

Note : (i) This question paper contains two Sections—Section A and Section B.

(ii) Both Sections are compulsory.

(iii) Answer any *two* sub-questions among (a), (b) and (c) in each main question of Section A. Each sub-question carries 10 marks.

(iv) Section B consisting of case study is compulsory. Section B is of 20 marks.

P. T. O.

Section—A

1. (a) Explain the need to mine data and what type of data can be mined. (CO1, 2)
- (b) Differentiate between supervised and unsupervised learning. Cite relevant examples. (CO1, 2)
- (c) Citing relevant examples explain business analytics process in detail. (CO1, 2)
2. (a) Explain hypothesis testing in detail also point out some reasons why hypothesis testing is important. (CO3)
- (b) Using the example of Amazon explain the significance of market basket analysis. (CO3)
- (c) Taking a real-life example explain why and how regression analysis is of utmost importance for a company. (CO3)
3. (a) Considering relevant example shed some light on properties of clustering. (CO4)
- (b) Using suitable examples assess the different type of outliers. (CO4)

- (c) "Clustering is an unsupervised machine learning-based algorithm that acts on unlabeled data." Critically analyze this statement. (CO4)
4. (a) Formulate the role and responsibilities of a data scientist. (CO5, 2)
- (b) Write short notes on data visualization. (CO5, 2)
- (c) Explain different phases of data analytics life cycle. (CO5, 2)

Section—B

5. **Case Study :** (CO5)

Unsupervised Machine Learning is the process of teaching a computer to use unlabeled, unclassified data and enabling the algorithm to operate on that data without supervision. Without any previous data training, the machine's job in this case is to organize unsorted data according to parallels, patterns, and variations.

The goal of clustering is to divide the population or set of data points into a number