

GUJARAT TECHNOLOGICAL UNIVERSITY
MCA– SEMESTER V- EXAMINATION –WINTER-2023

Subject Code:4659306**Date: 05/12/2023****Subject Name: Big Data Analytics****Time: 02:30 PM TO 05:00 PM****Total Marks: 70****Instructions:**

1. Attempt all questions.
2. Make Suitable assumptions wherever necessary.
3. Figures to the right indicate full marks.
4. Use of simple calculators and non-programmable scientific calculators are permitted.

- Q.1 (a) Define following. 04**
 1. Data warehouse
 2. Cluster centroid
 3. Stemming
 4. Term Frequency
- (b) Discuss the importance of analytic sandbox. 03**
(c) Explain Data preparation, Model planning and Model building phases of data analytic lifecycle. 07
- Q.2 (a) Explain types and drivers of big data. 07**
(b) Discuss collaborative filtering vs association rules. 07
- OR**
- (b) Discuss different metrics to measure the distance between two records. 07**
- Q.3 (a) Explain k-means clustering algorithm. 07**
(b) Write short note on item based collaborative filtering. 07
- OR**
- Q.3 (a) Discuss the concept of support and confidence for generating association rules in apriori algorithm. 07**
(b) Explain hierarchical clustering in brief. 07
- Q.4 (a) Explain simple exponential smoothing in detail. 07**
(b) Write short note on: A model with Linear Trend. 07
- OR**
- Q.4 (a) Define autocorrelation. Discuss ARIMA models in brief. 07**
(b) Explain edge list and adjacency matrix with proper example. 07
- Q.5 (a) Explain following: 07**
 1. Node level centrality metrics 2. Egocentric Network
(b) Explain advantages and disadvantages of social media analytics. 07
- OR**
- Q.5 (a) Explain tokenization in brief. 07**
(b) Explain bag-of-words approach with suitable example. 07
