

Seat No.: _____

Enrolment No. _____

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

Subject Code: 4659306

Date: 19/06/2023

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) i) What is “Bag-of-Words”? **03**
ii) What is BI? How it is different from data science? **03**
iii) What is Euclidean distance and Manhattan distance? **02**
(b) Discuss model planning and model building phases of data Analytics lifecycle.. **06**
- Q.2** (a) i) How to measure distance between two clusters? **04**
ii) What is the importance of selecting strong rules? **03**
(b) What is Big data? Discuss types of big data. **07**
- OR**
- (b) What is Data Preparation? Discuss importance of data preparation. **07**
- Q.3** (a) What is association rule mining? Discuss importance of association rule mining. **07**
(b) What is Collaborative filtering? Explain in brief various types of collaborative filtering. **07**
- OR**
- Q.3** (a) Explain the steps of k-Means Clustering Algorithm. **07**
(b) List various types of smoothing methods. Explain any one in detail using suitable example. **07**
- Q.4** (a) How to create ARIMA model for a time series. **07**
(b) Explain in brief moving average method of time series forecasting. **07**
- OR**
- Q.4** (a) How ARIMA models are powerful tool for forecasting time series data? **07**
(b) What is text mining? How text mining is different from data mining? **07**
- Q.5** (a) i) What is required in text mining? Why? **04**
ii) What is the importance of visualizing networks? **03**
(b) What is time series forecasting and regression based forecasting? **07**
- OR**
- Q.5** (a) What is descriptive and predictive time series forecasting? **07**
(b) What are the advantages and disadvantages of social network data collection? **07**
