

GUJARAT TECHNOLOGICAL UNIVERSITY**BACHELOR OF ENGINEERING – SEMESTER 7 – EXAMINATION – SUMMER 2026****Subject Code: 2170715****Date: 29/05/2026****Subject Name: Data Mining and Business Intelligence****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. Simple and non-programmable scientific calculators are allowed.**

- Q.1**
- (a) What is Business Intelligence (BI)? What is the role of data warehouse in achieving BI? **03**
- (b) Discuss the application of data warehousing and data mining. **04**
- (c) Explain KDD process in detail. **07**

- Q.2**
- (a) Explain star schema. **03**
- (b) Differentiate between OLAP and OLTP. **04**
- (c) Generate strong association rules for given dataset. (Min_supp = 60% and Min_conf = 80%) **07**

Team ID	Item_Bought
1	{M,O,N,K,E,Y}
2	{D,O,N,K,E,Y}
3	{M,A,K,E}
4	{M,U,C,K,Y}
5	{C,O,O,K,I,E}

OR

- (c) What is data cleaning? Explain Data Cleaning process for missing values & Noisy data. **07**

- Q.3**
- (a) Define support, confidence, and strong association. **03**
- (b) Explain Mean, Median, Mode and Variance in brief. **04**
- (c) Discuss importance of dimension reduction and algorithms used for it. **07**

OR

- Q.3**
- (a) Why market basket analysis is important? **03**
- (b) What is Cuboid? Explain various OLAP Operations on Data Cube in brief. **04**
- (c) List various data pre-processing tasks. Explain briefly any three tasks. **07**

- Q.4**
- (a) Explain Support and Confidence measures by taking appropriate example. **03**
- (b) Explain Bayesian learning and inference. **04**
- (c) What is Information Gain? Explain classification based on Decision Tree induction with the help of a suitable example. **07**

OR**Q.4**

- (a)** Write advantages of Business Intelligence. **03**
- (b)** Explain role of business intelligence in any one business domain. Discuss how it helps in business growth. **04**
- (c)** Explain linear and non-linear regression. **07**

- Q.5**
- (a)** Differentiate between classification and clustering. **03**
 - (b)** Difference between Data Mining and Text Mining **04**
 - (c)** Write a short note on Hadoop MapReduce. **07**

OR

- Q.5**
- (a)** Discuss fraud detection and click-stream analysis using data mining. **03**
 - (b)** What is Big Data? Discuss the 4 V's of Big Data. **04**
 - (c)** What is distributed file system? Explain HDFS architecture in detail. **07**
