

3 Hours)

Total Marks: 80

- N.B.: (1) Q. 1 is compulsory.
(2) Attempt any three out of the remaining five.
(3) Figures to the right indicate full marks.

80 Marks

- Q.1** A Differentiate DM and OLAP. 5
B Write a note on Web usage Mining 5
C Distinguish between parallel database and distributed database. 5
D Why data preprocessing is important in data mining? 5
- Q.2** A Define Parallel Database. Explain different types of parallel database architectures in detail. 10
B Define OLAP. Discuss OLAP models with a neat diagram. 10
- Q.3** A Define Classification. Explain KNN algorithm with suitable example. 10
B What is dimensional modeling? Discuss dimensional modelling techniques in data warehouse. 10
- Q.4** A What is agglomerative clustering? Explain with an example. 10
B Explain Data Reduction techniques in detail. 10
- Q.5** A 10

Word	Email Length	Attachment	Spam
Discount	Short	Picture	Yes
Discount	Short	Document	Yes
Hello	Long	Document	Yes
Thanks	Long	Document	Yes
Thanks	Long	Picture	No
Thanks	Short	Picture	Yes
Hello	Short	Document	No
Discount	Short	Picture	No
Discount	Long	Document	No
Thanks	Long	Document	No
Discount	Long	Picture	No
Hello	Short	Picture	Yes
Hello	Short	Picture	Yes
Thanks	Short	Document	Yes

Classify a new email which is a long email, has a picture and has the word discount in it using Naïve Bayes as spam or non spam.

- B Explain text mining and discuss in brief the Information retrieval methods 10

- Q.6** **A** What is association rule mining, and what is its primary goal? What is the Apriori algorithm. How does the Apriori algorithm handle the generation of frequent itemsets? **10**
- B** Define Abstract data type. Discuss the Operations on Structured Data. **10**
-