| Roll No | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|

# PRESIDENCY UNIVERSITY
# BENGALURU

## SCHOOL OF INFORMATION SCIENCE
## END TERM EXAMINATION - JUN 2023

**Semester :** Semester IV - 2021
**Course Code :** CSA2007
**Course Name :** Sem IV - CSA2007 - Data Mining
**Program :** BCA

**Date :** 12-JUN-2023
**Time :** 1.00PM - 4.00PM
**Max Marks :** 100
**Weightage :** 50%

**Instructions:**
*(i) Read all questions carefully and answer accordingly.*
*(ii) Question paper consists of 3 parts.*
*(iii) Scientific and non-programmable calculator are permitted.*
*(iv) Do not write any information on the question paper other than Roll Number.*

### PART A

**ANSWER ALL THE QUESTIONS** **(5 X 2 = 10M)**

**1.** What is Text Mining?
(CO5) [Knowledge]

**2.** What is overfitting and underfitting in classification?
(CO4) [Knowledge]

**3.** Write about Dimensionality reduction methods?
(CO2) [Knowledge]

**4.** Write any two differences between descriptive and predictive data mining tasks.
(CO1) [Knowledge]

**5.** What is Market Basket Analysis?
(CO3) [Knowledge]

### PART B

**ANSWER ALL THE QUESTIONS** **(5 X 10 = 50M)**

**6.** Write a short note on, i)Random Subsampling ii)k-fold cross validation.
(CO4) [Comprehension]

**7.** Why is outlier detection an important problem? Discuss the challenges involved in outlier detection.
(CO5) [Comprehension]

**8.** Identify Frequent Patterns for the given dataset {M,O,N,K,E,Y} {D,O,N,K,E,Y} {M,A,K,E} {M,U,C,K,Y} {C,O,O,K,I,E} using FP-growth algorithm. [Support= 60%, Confidence = 80%].

(CO3) [Comprehension]

**9.** What is Text Mining? Explain the different components of Text Mining.

(CO5) [Comprehension]

**10.** Expain Hierarchical Clustering in detail.

(CO4) [Comprehension]

## PART C

**ANSWER ALL THE QUESTIONS** **(2 X 20 = 40M)**

**11.** Construct decision tree using ID3 algorithm for the following data :

| Instance | Classification | a1 | a2 |
|---|---|---|---|
| 1 | + | T | T |
| 2 | + | T | T |
| 3 | - | T | F |
| 4 | + | F | F |
| 5 | - | F | T |
| 6 | - | F | T |

(CO4) [Application]

**12.** Apply FP Growth Algorithm for the given data set. Construct FP Tree and also generate Association Rules. [Min_Sup = 3, confidence=60%].

| Transaction ID | Items |
|---|---|
| T1 | {E,K,M,N,O,Y} |
| T2 | {D,E,K,N,O,Y} |
| T3 | {A,E,K,M} |
| T4 | {C,K,M,U,Y} |
| T5 | {C,E,I,K,O,O} |

(CO3) [Application]