

Roll No



**PRESIDENCY UNIVERSITY
BENGALURU**

SET B

**SCHOOL OF ENGINEERING
END TERM EXAMINATION - DEC 2023**

Semester : Semester VII - 2020

Course Code : CSE3134

Course Name : Text Mining and Text Analytics

Program : B.Tech.

Date : 05-JAN-2024

Time : 9:30AM - 12:30 PM

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

4 X 5M = 20M

1. Explain the probability of observing a word with an example
(CO3) [Knowledge]
2. Explain Syntagmatic Relation with an example.
(CO2) [Knowledge]
3. Describe the Landscape of Text Mining and Analytics
(CO1) [Knowledge]
4. Explain F-score evaluation.
(CO4) [Knowledge]

PART B

ANSWER ALL THE QUESTIONS

5 X 10M = 50M

5. Explain how topic extraction model will help to gather valuable information.
(CO3) [Comprehension]
6. Explain the similarity measure process in text mining can be used to identify the suitable clustering algorithm for a specific problem.
(CO3) [Comprehension]

7. Explain how PLSA model at a lower-stage, to estimate the parameters of models at a higher-stage to which few people belong.
(CO3) [Comprehension]
8. Explain how a conditional entropy-based feature selection method can help in overcome the limitation.
(CO3) [Comprehension]
9. Explain how the clustering task as a means of evaluating the ability of text representations to produce meaningful groups.
(CO3) [Comprehension]

PART C

ANSWER ALL THE QUESTIONS

2 X 15M = 30M

10. A group of 50 college students are given a self-administered questionnaire and asked how often they have used recreational drugs in the past year: Often (more than 5 times), Seldom (1 to 4 times), and Never (0 times). On another occasion, the same group of students was asked the same question in an interview. The following table shows their responses. Determine how closely their answers agree.

Interview	Questionnaire			Total
	Seldom	Often	Never	
Seldom	20	8	2	
Often	12	32	4	
Never	0	6	16	
Total				

Determine how closely their answers agree.

(CO2) [Application]

11. Let C1 and C2 be two coins.
 \emptyset_1 be the probability of getting head with C1.
 \emptyset_2 be the probability of getting head with C2.
 Chosing any of the coin randomly, toss for 5 times.
 Each selected coin has to toss for 10 mins.

C2	H	T	T	T	H	H	T	H	T	H
C1	H	H	H	H	T	H	H	H	H	H
C1	H	T	H	H	H	H	H	T	H	H
C2	H	T	H	T	T	T	H	H	T	T
C1	T	H	H	H	T	H	H	H	T	H

Find value of \emptyset_1 and \emptyset_2 by tossing C1 and C2 for 10 times by assuming the probabilities $\emptyset_1 = 0.6$ and $\emptyset_2 = 0.5$

(CO4,CO3) [Application]