

Name: Enrolment No:	
--------------------------------------	---

UNIVERSITY OF PETROLEUM AND ENERGY STUDIES

End Semester Examination, May 2022

Course: Text Mining
Program: BBA Analytics & Big Data
Course Code: DSBA 3005B

Semester: 6
Time : 03 hrs.
Max. Marks: 100

Instructions: *The student must write his/her name and enrolment no. in the space designated above.*

SECTION A
10Qx2M=20Marks

S. No.		Marks	CO
Q1	_____ is the process of transforming unstructured text into a structured format to identify meaningful patterns and new insights. a. Data mining b. Text mining c. File mining d. Deep mining	2	CO1
Q2	Text mining is being used by large media companies, to clarify information and to provide readers with greater search experiences, a. TRUE b. FALSE c. Can be true or false d. Can not say	2	CO1
Q3	Stemming: This refers to the process of separating the prefixes and suffixes from words to derive the root word form and meaning. a. TRUE b. FALSE c. Can be true or false d. Can not say	2	CO1
Q4	Text mining reads an _____ form of data to provide meaningful information patterns a. Structured b. Unstructured c. Semi-Structured d. None of the above.	2	CO1
Q5	_____ is the application of data mining techniques to discover patterns from the Web a. Text Mining b. Multimedia Mining c. Web Mining d. Link Mining	2	CO1
Q6	In text mining, inputs to the process include unstructured data such as Word documents, PDF files, text excerpts, e-mail and XML files	2	CO1

	(True/False)		
Q7	NLP (Natural Language Processing) is concerned with the interactions between computers and human languages. (True/False)	2	CO1
Q8	Finding out the most mentioned words in unstructured text can be particularly useful when analyzing: a. customer reviews b. social media conversations c. customer feedback. d. All the above	2	CO1
Q9	The process of breaking down a text paragraph into smaller chunks such as words or sentence is called _____.	2	CO1
Q10	Text may contain _____ such as is, am, are, this, a, an, the, etc which are considered as noise in the text.	2	CO1
SECTION B 4Qx5M= 20 Marks			
Q1	In the area of Sentiment analysis, how is ‘Stemming’ process useful?	5	CO1
Q2	What are ‘Stop Words’? Why are they not of significance?	5	CO1
Q3	What is the most fundamental use of ‘Tokenization’ in the area of Text Mining?	5	CO2
Q4	How does Text Mining help in ‘Spam detection’?	5	CO1
SECTION-C 3Qx10M=30 Marks			
Q1	Differentiate between Text Mining, Text Analysis, and Text Analytics.	10	CO2
Q2	How does Sentiment Analysis help businesses?	10	CO2
Q3	Give examples where Machine Learning is helping Businesses and customers. OR Elucidate upon the disadvantages of Text Mining.	10	CO2
SECTION-D 2Qx15M= 30 Marks			
Q1	Discuss some basic methods employed in Text Mining.	15	CO3
Q2	List down the various industrial applications of Text Mining. OR Indicate the key differences between Text Mining & Sentiment Analysis.	15	CO3