

**K. J. Somaiya Institute of Technology, Sion, Mumbai-22**  
(Autonomous College Affiliated to University of Mumbai)

May 2024		I	
(B.Tech.) Program: DS/AIML Scheme			
Examination: Honours		Semester: VIII	
Course Code: <b>HDSC801</b>	Course Name: <b>Text, Web and Social Media Analytics</b>		Max. Marks: 60
Date of Exam: <b>24/05/2024</b>	Duration: 2.5 Hours		

**Instructions:**

- (1) All questions are compulsory.
- (2) Draw neat diagrams wherever applicable.
- (3) Assume suitable data, if necessary.

		Max. Marks	CO	BT level
<b>Q 1</b>	<b>Solve any six questions out of eight.</b>	<b>12</b>		
i)	Explain 'Information Extraction' from text.	2	CO1	U
ii)	Explain 'Group Spam Detection'	2	CO6	U
iii)	Explain 'Web Usage Mining' in detail.	2	CO3	U
iv)	Explain Item-based CF.	2	CO5	U
v)	Explain 'Types of data' in web usage mining.	2	CO4	U
vi)	Calculate the Golomb encoding of a suitable number when the base b is power of two.	2	CO3	Ap
vii)	What is Entropy? The dataset has 9 positive instances and 5 negative instances, therefore calculates the entropy of this data set	2	CO2	Ap
viii)	Explain 'Non-Negative Matrix Factorization (NMF)'.	2	CO2	U
<b>Q.2</b>	<b>Solve any four questions out of six.</b>	<b>16</b>		
i)	Explain relation extraction w.r.t. information extraction	4	CO1	U
ii)	Explain 'Hidden Markov Model (HMM)' in detail with its advantages and disadvantages.	4	CO2	U
iii)	Apply the 'Elias Delta' and 'Elias Gamma' compression technique on following given numbers and create their coding and decoding forms separately? a) 6 b)7 c) 8 d) 10	4	CO3	Ap
iv)	Explain the 'Problem of Opinion Mining' in detail with various examples.	4	CO6	U

**K. J. Somaiya Institute of Technology, Sion, Mumbai-22  
(Autonomous College Affiliated to University of Mumbai)**

May 2024		
(B.Tech.) Program: <b>DS/AIML</b> Scheme <b>T</b>		
Examination: <b>Honours</b> Semester: <b>VIII</b>		
Course Code: <b>HDSC801</b>	Course Name: <b>Text, Web and Social Media Analytics</b>	Max. Marks: 60
Date of Exam:	Duration: 2.5 Hours	

v)	What is the main idea behind the recommender system? Explain the challenges of Recommender systems.	4	CO5	U
vi)	Explain 'Analysis of Sequential and Navigational Patterns' for web usage mining.	4	CO4	U
<b>Q.3</b>	<b>Solve any two questions out of three.</b>	<b>16</b>		
i)	Explain 'Data Pre-Processing Elements' of web usage mining in detail.	8	CO4	U
ii)	Explain in detail about 'Web Search', also illustrate in detail about 'Meta-Search'.	8	CO3	U
iii)	Explain N-Gram Model in detail with an example. Also discuss the applications and limitations of the N-gram Model in NLP.	8	CO1	U
<b>Q.4</b>	<b>Solve any two questions out of three.</b>	<b>16</b>		
i)	What is Document Sentiment Classification? Apply the Unsupervised classification learning technique on one of the example to describe its three steps.	8	CO6	Ap
ii)	Solve and calculate user's rating for a particular item on specific data using the 'Item-Based Collaborative Filtering' method. ( Assume that neighborhood size = 2 )	8	CO5	Ap
iii)	What is Probabilistic-Document clustering? Use Topic Model (LDA) concept to describe it.	8	CO2	Ap

\*\*\*\*\*