Honours Program Sem - VIII

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

AIML

July-Aug 2025
~~May-June 2025~~

(**B.Tech.**) Program: ~~**DS**~~**/AIML** Scheme :**IIB**

Supplementary Examination: **LY- Honours** Semester: **VIII**

Course Code: **HDSC801** Course Name: **Text, Web and Social Media Analytics**

Date of Exam: ~~28/05/25~~ 25/07/25 Duration: 2.5 Hours Max. Marks: 60

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

Q. No.	Question	Max. Marks	CO	BT level
Q 1	Solve any **two** questions out of three: (05 marks each)	10		
a)	Give a detailed explanation of the **N-Gram** model using an example. Additionally, describe its applications and limitations in NLP.		CO1	U
b)	Discuss **LDA** probabilistic modeling algorithm in detail.		CO2	U
c)	Explain '**Inverted Index**' of a document collection using an example.		CO3	U
Q 2	Solve any **two** questions out of three: (05 marks each)	10		
a)	Explain with examples '**Node Neighborhood-Based Methods**' in detail		CO4	U
b)	Use any of the learning strategies for classification of sentiments present in the document.		CO5	Ap
c)	Explain discovery and analysis of '**Web Usage Patterns**'.		CO3	U
Q.3	Solve any **two** questions out of three. (10 marks each)	20		
a)	Discuss in detail on '**Unsupervised Information Extraction**' and '**Kernel**' methods' for relation extraction.		CO1	U
b)	Discuss in detail about '**Opinion Lexicon Expansion**'.		CO5	U
c)	With the example of modeling navigational trails, describe the analysis of sequential and navigational patterns in detail for web usage mining.		CO3	U
Q.4	Solve any **two** questions out of three. (10 marks each)	20		
a)	Apply the item-based collaborative filtering on following data (User-Item Rating Matrix), calculate the User **U1's rating for Item E**.		CO4	Ap

Seat No.:

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

~~May-June 2025~~ July - Aug 2025

(**B.Tech.**) Program: **DS/AIML** Scheme :**IIB**

~~Supplementry~~ Examination: **LY- Honours** Semester: **VIII**

Course Code: **HDSC801** Course Name: **Text, Web and Social Media Analytics**

Date of Exam: ~~28/05/25~~ 25/07/2025 Duration: 2.5 Hours Max. Marks: 60

		CO2	Ap
b)	I) Discuss the '**Latent Dirichlet Allocation**' topic model in detail. [5] II) A sample of 1000 patients from a healthcare database with two binary indicators: Hypertension (H): Yes or No & Smoker (S): Yes or No [5] A breakdown of the data is : i) Find the probability that a patient has hypertension and is a smoker. ii) Probability that a patient has hypertension. iii) Probability that a patient is a smoker. iv) Calculate the percentage(%) of smokers who have hypertension		
c)	Explain Index Compression in detail. Apply the **Golomb Encoding** compression technique on following given numbers and create their coding forms separately. i) Encode a number (n)= 12 with base = 4 ii) Encode a number (n)= 23 with base = 7 iii) Encode a number (n)= 03 with base = 5	CO3	Ap

User	Item A	Item B	Item C	Item D	Item E
U1	5	3	4	4	?
U2	3	1	2	3	3
U3	4	3	4	3	5
U4	3	3	1	5	4
U5	1	5	5	2	1

	Smoker(Yes)	Smoker(No)	Total
Hypertension (Yes)	180	220	400
Hypertension (No)	120	480	600
Total	300	700	1000

****

Seat No.: