

NIRMA UNIVERSITY

Institute:	Institute of Technology
Name of Programme:	Integrated B.Tech.(CSE)-MBA
Course Code:	CSI0909
Course Title:	Text Analytics
Course Type:	(<input type="checkbox"/> Core/ <input type="checkbox"/> Value Added Course / <input checked="" type="checkbox"/> Department Elective / <input type="checkbox"/> Institute Elective/ <input type="checkbox"/> University Elective/ <input type="checkbox"/> Open Elective / <input type="checkbox"/> Any other)
Year of Introduction:	2022-23

L	T	Practical Component				C
		LPW	PW	W	S	
3	-	2	-	-	-	4

Course Learning Outcomes (CLOs):

At the end of the course, the student will be able to –

1. demonstrate quantitative analysis of text data (BL2)
2. apply standard methods including machine learning techniques for analyzing and retrieving information from text data (BL3)
3. evaluate performance of text processing algorithms used for text information retrieval and analysis (BL5)
4. interpret the results and recommend possible actions from analytics performed on text data (BL5)

Syllabus:

Total Teaching hours: 30

Unit	Syllabus	Teaching hours
Unit-I	Introduction to Text Analytics: Overview of text data and text data analytics, applications and recent tools for performing text analytics	03
Unit-II	Information Retrieval Systems: Text Preprocessing, Bag of words model, Vector space model, Document representation using boolean, TF and TF-IDF schmes, basic query processing, evaluating IR systems using precision, recall, and P-R curve, relevance feedback, similarity measures, meta search techniques	10
Unit-III	Text Analytics Framework: Revisiting Bayes theorem and naive Bayesian classification on text data, Markov models and language models, n-grams, hidden Markov models and sequence models	07
Unit-IV	Advanced topics: Recommender systems, Sentiment analysis, Spam email classification, Question-answering systems.	10

Self-Study: The self-study contents will be declared at the commencement of semester. Around 10% of the questions will be asked from self-study contents.

Suggested Readings/References:

1. Manning, C. D., Raghavan, P., and Schütze, H., Introduction to Information Retrieval. Cambridge University Press
2. Miner, G., Delen, D., Elder, J., Fast, A., Hill, T., and Nisbet, A. R., Practical Text Mining and Statistical Analysis for Non-structured Text Data Applications, Elsevier Inc.
3. Bird, S., Klein, E., & Loper, E., Natural Language Processing with Python. O'Reilly Media, Inc.
4. B. Croft, D. Metzler, T. Strohman, Information Retrieval in Practice, Pearson Education

Suggested List of Experiments:

Sr. No.	Title	Hours
1	Implement different text preprocessing techniques on some standard text corpus.	02
2	Represent a text-corpus in the form of a Boolean document-term matrix. Perform query processing using logical operations.	02
3	Represent a text-corpus in the form of TF and TF-IDF based document-term matrix. Visualize the corpus using word cloud and histogram. Perform search using cosine similarity-based approach.	02
4	Design a GUI based mini search engine to perform the above tasks.	02
5	Implement meta search techniques on a given data.	02
6	Apply multinomial and Bernoulli naive Bayesian classifiers on text data.	02
7	Implement language model based on given text corpus.	02
8	Implement sequence model on a given text corpus.	02
9	Implement a text-based recommender system.	02
10	Implement sentiment analysis or emotion detection system based on text data.	02

Suggested Case List:

-NA-

