Course Code: CSE4712	Course Name: Biomedical Text Mining	
Credits: 3 (3-0-0)	**Contact Hours:** 3 hours Theory per week	
Batch: 2022, 7th Sem **Academic Year:** 2025-26	**Semester Duration:** 28th July 2025 to 28th Nov 2025	BMU BML Munjal University
Course Faculty: Yogesh Gupta 9068039711, Yogesh.gupta@bmu.edu.in	**Course Coordinator:** **Name: Yogesh Gupta** **Email:** Yogesh.gupta@bmu.edu.in **Office:** 28, IV Floor, E2 Building	

Aim of the course: Biomedical Text Mining is an interdisciplinary course that equips students with the skills to extract, analyze, and interpret information from unstructured biomedical literature using natural language processing (NLP) and machine learning techniques. The course explores how automated systems can process the vast and continuously growing biomedical data found in research articles, clinical reports, drug labels, and electronic health records.

Students will learn key concepts in information retrieval, named entity recognition, relation extraction, ontology-based annotation, and literature-based discovery specific to the biomedical domain. Practical sessions will involve applying biomedical NLP tools, working with corpora such as PubMed and ClinicalTrials.gov, and building simple biomedical text mining pipelines.

Course Overview and Context: With the exponential increase in biomedical publications and clinical data, manual curation is no longer feasible. Biomedical text mining provides scalable solutions for knowledge discovery, hypothesis generation, and clinical decision support. This course is highly relevant for students pursuing careers or research in bioinformatics, computational biology, biomedical data science, and health informatics.

It integrates concepts from computational linguistics, data science, and biomedical sciences, fostering interdisciplinary thinking and practical problem-solving skills. The course also prepares students to contribute to real-world challenges such as drug repurposing, adverse event detection, precision medicine, and systematic reviews.

Course Outcomes (CO): At the end of the course the students should be able to do the following:

	Course Outcome	Mapped Component	Activities
CO 1	Apply foundational biomedical text mining techniques, including biomedical terminologies, ontologies, and preprocessing of biomedical literature.	Skill Development, Employability	Tutorials using UMLS, PubMed data preprocessing, hands-on lab sessions with BioNLP tools
CO 2	Implement NLP and machine learning algorithms to extract meaningful insights from biomedical corpora such as clinical notes, articles, and reports.	Employability, Skill Development	Model building using Python (spaCy, BioBERT), real-world dataset analysis, ML model evaluation assignments
CO 3	Design innovative text mining applications for healthcare or pharmaceutical domains addressing unmet needs or gaps.	Entrepreneurship, Employability	Mini-projects, startup idea pitching, building prototypes for medical text-based search, chatbot, or alert system

Prerequisite: if any
1. Natural language Processing
2. Machine Learning

Topics of the course:

Topics	Number of sessions	Course Outcomes
Text Mining and Biomedical Data: Introduction	3	CO1
Biomedical Data Types, Sources, Content, and Retrieval.	2	CO1
Information Analysis Using Biomedical Text Mining.	3	CO1
Biomedical Data Visualization, Biomedical Text Data Visualization.	4	CO1, CO2
Biomedical Ontology and Model Building: Role of Ontology in Biomedical Text Mining, Ontology in Text Mining and Matching,	5	CO1, CO2
Fundamentals of Vector-Based Text Representation and Word Embeddings, Transformer-Based Models for Text Representation and Processing.	4	CO1, CO2
Tasks in Biomedical Text Mining: Information Retrieval and Query Expansion for Biomedical Data,	5	CO2
Advances in Biomedical Entity and Relation Extraction: Techniques and Applications,	4	CO2
Biomedical Document Clustering.	3	CO2
Knowledge Graph for Biomedical Text Mining: Exploring Knowledge Graphs (KG): A Comprehensive Overview.	5	CO2, CO3
Applications of Biomedical Text Mining: Text Mining for Telemedicine,	4	CO2, CO3
Text Mining for Recommendation Systems/Expert Systems in Health Domain,	4	CO3
Ethical Issues in Biomedical Text Mining,	2	CO3

CO/PO Mapping:

CO/PO Mapping	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3	PSO4
CO1	3	1				1							1			
CO2	2	2		1	1	1										
CO3	2		3	1	2	1	1	2			2	2	1		2	

Learning Resources:

Textbook:

1. Sophia Ananiadou and John McNaught (2021). Text Mining for Biology And Biomedicine. Kindle Edition.
2. Aditi Sharan, Nidhi Malik, Hazra Imran, Indira Ghosh (2024). Text Mining Approaches for Biomedical Data, Springer.

Reference Books:

1. Jonathan Pevsner. (2022). Bioinformatics and Functional Genomics. 3rd Edition.

Research papers and Articles:

1. Uzuner, Ö., et al. (2008). "Evaluating the state of the art in automatic de-identification."
2. Wei, C. H., et al. (2016). "tmVar 2.0: Integrating genomic variant information from literature with dbSNP and ClinVar for precision medicine."
3. Liu, H., et al. (2012). "BioMedLatent: A text mining tool for biomedical literature."

Online Platforms & MOOCs:

Coursera / edX / FutureLearn – Courses like:

- *Text Mining and Analytics* (University of Illinois)

- *Biomedical Data Science Specialization* (Johns Hopkins University)

- *Natural Language Processing Specialization* (DeepLearning.AI)

Datasets:

- PubMed Central (PMC) – Open-access full-text biomedical articles
- MIMIC-III – Clinical database for healthcare research
- GENIA Corpus – Biomedical abstracts annotated with linguistic information
- BioCreative Challenges – Datasets from biomedical NLP competitions
- BC5CDR – For chemical and disease recognition from text

Experiential Learning Component:

The Experiential Learning Component of Biomedical Text Mining is designed to provide students with hands-on experiences that reinforce theoretical knowledge through practical application. Throughout the course, students will engage in project-based learning where they will be assigned practical projects that evolve from Python, spaCy, scispaCy, BioBERT, MetaMap, NLTK, Hugging Face Transformers. These projects, which span the entirety of the course, are structured into phase-wise evaluations to ensure consistent application and progression of skills.

Assessment Pattern: The final grade will be determined by the marks or grades earned during the project's phase-wise evaluations and the end-term assessment. Grading will be conducted using the relative grading method outlined in the university's academic regulations. To be eligible for grading, students must achieve a minimum of 40% of the total marks upon completing all assessments listed in the table below:

Evaluation Component	Weightage (%)	Evaluation Schedule/Timeline	Rubrics/Remarks
Project	70%	Will be held throughout the Semester	**Focus:** Knowledge and application of Biomedical Text Mining. Mid Term (1st Evaluation): 30% End Term (Final Evaluation including report): 40%
Presentation	10%	Throughout the semester	Presentations will be on various topics covering applications of Biomedical text Mining.
Two Quiz	20%	September 2nd Week; November 1st Week	Topics to be covered will be announced in the class. 10 MCQ questions, each of 1 mark. Mode of quiz will be offline.

Student Responsibilities:

– Attend lectures and do the work Assignments as per instructions.
– Participate in the discussions/assignments held during classes.
– Check announcements on the LMS and emails regularly.
– Submit the assigned task on time.
– Regularly check marks on the LMS to ensure they are up to date.
– Participate in class and take necessary actions to grasp the material. Asking questions is encouraged.
– Communicate any concerns by speaking directly with the instructor.

Attendance Policy: Students are expected to attend classes regularly. Failure to follow the classes regularly and adhere to the expected attendance percentage will result in losing quiz/lab marks and a reduction of the grade as per the University's grading policy.

Recourse Examination Policy: In case a student fails the course, a one-time recourse is permitted as per the academic regulations of the University. Recourse is allowed only for the End Semester examination.

Make-up policy: No make-up exam will be conducted for unexcused absences. The faculty needs to be informed in advance in case the student is not going to appear for any evaluation component, and it is at the discretion of the faculty to sanction makeup for an evaluation component.

Behavior Expectations: No mobile phones and other distractive gadgets are permitted in the class.

Academic Dishonesty/Cheating/Plagiarism: Plagiarism and dishonesty in any form in any evaluation component will lead to appropriate disciplinary action.