

COURSE TITLE	NATURAL LANGUAGE PROCESSING
COURSE CODE	05MD0307
COURSE CREDITS	5

Course Outcomes: After completion of this course, student will be able to:

- 1 Understand the basics of NLP its applications and challenges in data science.
- 2 Develop the techniques for text classification and language modeling.
- 3 Design and implement NLP models for tasks such as text sentiment analysis and information retrieval.
- 4 Apply NLP techniques to real-world applications, such as machine translation, and text representation.
- 5 Develop the ethical and social considerations in NLP applications.

Pre-requisite of course: Modern machine learning algorithms and involves a lot of math and programming.

Teaching and Examination Scheme

Theory Hours	Tutorial Hours	Practical Hours	ESE	IA	CSE	Viva	Term Work
4	1	0	50	30	20	0	0

Contents : Unit	Topics	Contact Hours
1	Introduction to NLP Overview of NLP and its applications, Basic concepts and challenges in NLP, Tokenization, NLP Pipeline and components, Stop word removal and filtering, Stemming case conversion and Lemmatization, Handling Spelling errors and noise in text data	10
2	Language Modeling and Text Classification N-gram language models, POS (Part-of-Speech) tagging, Named Entity Recognition (NER), Supervised learning algorithms for text classification (e.g., Naive Bayes, SVM), Feature extraction for text classification	10
3	Sentiment Analysis and Information Reterival Opinion mining, Lexicon-based approaches, Machine learning approaches, Aspect-based sentiment analysis, Information Retrieval Basics, Retrieval models: Vector Space Model, BM25, Evaluation metrics for information retrieval, Search engines and query processing, Text indexing and relevance ranking	10

Contents : Unit	Topics	Contact Hours
4	Ethical and Social Considerations in NLP, NLP and Deep Learning Bias and fairness in NLP systems, Privacy and security issues, Responsible AI practices in NLP, Recurrent Neural Networks (RNNs) for sequence modeling, Attention mechanisms for NLP tasks, Transformer models (e.g., BERT, GPT) for NLP applications	10
5	Machine Translation and Text Representation Statistical machine translation, Neural machine translation (e.g., sequence-to-sequence models), Text generation using language models, Rule-based machine translation, Evaluation metrics for machine translation, Bag-of-Words model, TF-IDF (Term Frequency-Inverse Document Frequency), Word embeddings: Word2Vec, GloVe	10
Total Hours		50

Textbook :

- 1 Speech and Language Processing, Daniel Jurafsky, James H. Martin, PHI, 3rdE

References:

- 1 Foundations of Statistical Natural Language Processing, Foundations of Statistical Natural Language Processing, Christopher D. Manning and Hinrich Schütze., ., .
- 2 Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit, Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit, Steven Bird, Ewan Klein, and Edward Loper, ., .
- 3 Deep Learning for Natural Language Processing, Deep Learning for Natural Language Processing, Palash Goyal, Sumit Pandey, and Karan Jain., ., .

Suggested Theory Distribution:

The suggested theory distribution as per Bloom's taxonomy is as follows. This distribution serves as guidelines for teachers and students to achieve effective teaching-learning process

Distribution of Theory for course delivery and evaluation					
Remember / Knowledge	Understand	Apply	Analyze	Evaluate	Higher order Thinking
10.00	10.00	30.00	20.00	20.00	10.00

Instructional Method:

- 1 Boardwork
- 2 PPT
- 3 Videos

Supplementary Resources:

- 1 <https://www.javatpoint.com/nlp>
- 2 <https://www.kaggle.com/learn/natural-language-processing>