

PRASAD V. POTLURI SIDDHARTHA INSTITUTE OF TECHNOLOGY

(Autonomous) Kanuru, Vijayawada-520007

Department of Computer Science & Engineering (Data Science)

IV B. Tech – I Semester

Natural Language Processing

Course Code	23DS4701D	Year	IV	Semester	I
Course Category	PCC	Branch	CSE (DS)	Course Type	Theory
Credits	3	L-T-P	3-0-0	Prerequisites	Machine Learning
Continuous Internal Evaluation	30	Semester End Examination	70	Total Marks	100

Course Outcomes

Upon successful completion of the course, the student will be able to

CO1	Describe the fundamental concepts, techniques, and applications of NLP to understand how machines process and interpret text.	L2
CO2	Apply various text pre-processing techniques such as tokenization, normalization, stemming, and lemmatization to prepare clean and structured text data suitable for NLP tasks.	L3
CO3	Use pre-trained language models such as BERT and GPT on task-specific datasets to achieve effective performance in downstream NLP applications.	L3
CO4	Analyze the performance and behavior of NLP models using evaluation metrics to identify strengths and limitations.	L4

Contribution of Course Outcomes towards achievement of Program Outcomes & Strength of correlations (3: High,2: Medium, 1: Low)

[illegible]

Syllabus		
Unit No.	Contents	Mapped CO
I	Natural Language Processing: Introduction, History and evolution, Phases in NLP, Applications of NLP Basic Text Processing: Text representation (Words, sentences, and documents), Basic text processing (Tokenization, stemming, lemmatization), Language models: Understanding syntax, semantics, and pragmatics of language, Unigrams, bigrams, trigrams	CO1, CO2
II	Text Preprocessing: Cleaning and normalizing text, Removing stop words, punctuation, special characters, Handling misspellings and text normalization, Sentence segmentation and word segmentation Feature Engineering for NLP: N-grams, Part-of-Speech (POS) tagging, Named Entity Recognition (NER), Bag of Words (BoW) model, Term Frequency-Inverse Document Frequency (TF-IDF), Word embedding, Evaluating and visualizing text features	CO1, CO2
III	Bidirectional Encoder Representations from Transformers (BERT): Introduction, significance in NLP, BERT architecture, Pre-training and fine-tuning phases of BERT, applications of BERT.	CO1, CO3, CO4
IV	Transformer: Introduction, Transformer Architecture, Self-Attention, Multi- Head Attention, Encoder-decoder architecture. Transformer-based models: Training techniques for Transformer models, applications of Transformers.	CO1, CO3, CO4
V	Generative Pre-trained Transformer: Introduction, Training process of GPT models, Fine-tuning GPT for Downstream Tasks, Ethical Considerations and Bias in NLP, Applications of GPT.	CO1, CO3, CO4
Learning Resources		
Text Books		
1. Deep Learning for Natural Language Processing, Palash Goyal, Sumit Pandey, Karan Jain, First Edition, 2018, Apress. 2. Speech and Language Processing, Daniel Jurafsky and James H.Martin, Third Edition, 2024.		
References		
1. Statistical Language Learning, Charniack, Eugene, 1993, MIT Press. 2. Foundations of Statistical Natural Language Processing, Christopher Manning, Schutze Heinrich, 1999, MIT Press. 3. Natural Language Understanding, James Allen, 2003, Pearson Education.		
e- Resources and other Digital Material		
1. https://web.stanford.edu/class/cs224n/ 2. https://web.stanford.edu/~jurafsky/slp3/ed3bookaug20_2024.pdf 3. https://github.com/Denis2054/Transformers-for-NLP-and-Computer-Vision-3rd-Edition		