PRASAD V. POTLURI SIDDHARTHA INSTITUTE OF TECHNOLOGY

(Autonomous) Kanuru, Vijayawada-520007

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING (AI & ML)

III B. Tech – II Semester

Natural Language Processing

Course Code	20AM3603	**Year**	III	**Semester**	II
Course Category	PCC	**Branch**	CSE (AI&ML)	**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 Natural Language Processing (NLP)	**L2**
CO2	Apply various text pre-processing techniques to prepare text data in a suitable format for NLP tasks.	**L3**
CO3	Apply pre-trained language models like BERT and GPT for downstream tasks by fine-tuning these models on specific datasets	**L3**
CO4	Analyze the performance and behaviour of NLP models in various NLP tasks.	**L4**

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

	PO1	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**	**PO7**	**PO8**	**PO9**	**PO10**	**PO11**	**PO12**	**PSO1**	**PSO2**
CO1	2													
CO2	3											2	3	
CO3	3											2	3	
CO4		3										2		

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**
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, bidirectional context, 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.
3. Transformers for Natural Language Processing and Computer Vision, Denis Rothman , 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