

# 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 Lab

<b>Course Code</b>	20AM3652	<b>Year</b>	III	<b>Semester</b>	II
<b>Course Category</b>	PCC Lab	<b>Branch</b>	CSE (AI&ML)	<b>Course Type</b>	Practical
<b>Credits</b>	1.5	<b>L-T-P</b>	0-0-3	<b>Prerequisites</b>	Python Programming
<b>Continuous Internal Evaluation</b>	15	<b>Semester End Examination</b>	35	<b>Total Marks</b>	50

Course Outcomes		
Upon successful completion of the course, the student will be able to		
<b>CO1</b>	Demonstrate experimental procedures through oral communication and submit comprehensive documentation reports.	<b>L2</b>
<b>CO2</b>	Implement various NLP pre-processing techniques and deep learning models to solve real-world text processing tasks using tools.	<b>L3</b>
<b>CO3</b>	Analyze the performance and behavior of different NLP models on various tasks, identifying strengths and weaknesses.	<b>L4</b>
<b>CO4</b>	Critically evaluate the strengths and limitations of the deep learning models used in the experiments, considering factors like accuracy, robustness, and potential biases.	<b>L5</b>

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
<b>CO1</b>	2									2				
<b>CO2</b>					3							3	2	
<b>CO3</b>		3										3		
<b>CO4</b>				3								3		

Syllabus		
Expt. No.	Contents	Mapped CO
1.	Explore different Tools: nltk, spaCy, MonkeyLearn, MindMeld, OpenAI, Gensim	CO1
2.	Preprocess text data by performing text cleaning and normalization techniques, including removing stop words, punctuation marks, special characters, and converting the text to lowercase or uppercase as required.	CO1 to CO4
3.	Implement a text preprocessing pipeline for a given dataset that includes tokenization, and stemming/lemmatization.	CO1 to CO4
4.	Implement Bag-of-Words (BoW) and Term Frequency-Inverse Document Frequency (TF-IDF) vectorization techniques for transforming textual data into numerical feature representations suitable for machine learning models.	CO1 to CO4
5.	Train and visualize word embeddings using Word2Vec and GloVe algorithms.	CO1 to CO4
6.	Implement attention mechanisms in Recurrent Neural Networks (RNNs)	CO1 to CO4
7.	Implement and Fine-tuning Bidirectional Encoder Representations from Transformers (BERT) for Named Entity Recognition (NER) on a Specific Domain	CO1 to CO4
8.	Implement and fine-tune a Transformer model for text classification.	CO1 to CO4
9.	Implement text generation using GPT	CO1 to CO4
10.	<b>Capstone Project-1: Sentiment Analysis of Product Reviews</b> <ul style="list-style-type: none"> <li>Utilize pre-trained language models like BERT or GPT to build a sentiment analysis system</li> <li>Train and fine-tune the model on a dataset of product reviews</li> <li>Explore techniques for handling imbalanced data and improving sentiment classification performance</li> </ul>	CO1 to CO4
11.	<b>Capstone Project-2: Question Answering System for Customer Support</b> <ul style="list-style-type: none"> <li>Fine-tune a pre-trained BERT or GPT model for question answering</li> <li>Train the model on a dataset of customer queries and corresponding answers</li> <li>Deploy the question answering system for customer support applications</li> </ul>	CO1 to CO4
12.	<b>Capstone Project-3: Conversational AI for Chatbots</b> <ul style="list-style-type: none"> <li>Develop a conversational AI system using pre-trained language models and attention mechanisms</li> <li>Train the model on dialogue datasets</li> <li>Explore techniques for handling context and maintaining coherent conversations</li> </ul>	CO1 to CO4

## Learning Resources

### Text Books and Reference Books

1. Natural Language Processing with Python – Analyzing Text with the Natural Language, Toolkit Steven Bird, Ewan Klein, and Edward Loper, 1st edition, 2009, O'Reilly Media.
2. Practical Natural Language Processing: A Comprehensive Guide to Build Real World NLP Systems, Sowmya Vajjala, Bodhisattwa Majumder, Anuj Gupta, Harshit Surana, Greyscale Indian Edition, 1st edition, 2020, O'Reilly Media.

### e-Resources & other digital material

1. <http://www.nptelvideos.in/2012/11/natural-languageprocessing.html>
2. <https://www.nltk.org/>
3. <https://huggingface.co/docs/transformers/en/index>
4. <https://nlpprogress.com/>