NIRMA UNIVERSITY

Institute:	Institute of Technology, School of Technology
Name of Programme:	MTech CSE, MTech CSE (Data Science)
Course Code:	6CS361ME25
Course Title:	Natural Language Computing
Course Type:	Department Elective-I
Year of Introduction:	2025-26

L	T	Practical Component				C
		LPW	PW	W	S	
3	0	2	-	-	-	4

Course Learning Outcomes (CLO):

At the end of the course, students will be able to –

1. demonstrate foundational concepts in natural language processing and neural network models (BL2)
2. analyse and implement vector space models for word representation (BL4)
3. develop recurrent neural networks and attention mechanisms for advanced NLP tasks (BL6
4. adapt transformer models for diverse NLP applications, addressing ethical and interpretability aspects and societal impacts. (BL6)

Unit	Contents	Teaching Hours (Total 45)
Unit-I	**Introduction to NLC:** Overview of Natural Language Processing, Text Pre-processing, N-gram Models, NLP and Neural Networks, Biological and Artificial Neural Networks, Perceptron Learning, Linear Separability, Feed Forward Neural Networks, Gradient Descent and Backpropagation	06
Unit-II	**Vector Space Models:** Word2Vec – Skiagram and Continuous Bag of Words Model, Glove, Doc2Vec	04
Unit-III	**Recurrent Neural Networks and Attention Mechanisms:** Simple RNN, RNN Architectures, GRU, LSTM, Bi-Directional and Deep RNNs, Seq2Seq Models and Machine Translation, Simple Dot-product Attention, Multiplicative and Additive Attention, Language Modelling	07
Unit-IV	**Transformer Models:** Self-Attention, Encoder Transformer, Decoder Transformer, Encoder-Decoder Transformer, Pretraining, Post-training (Prompting, RLHF, SFT, DPO), Adding Knowledge to Language Models, BERT, GPT, T5	12
Unit-V	**Interpretability and Fine-tuning:** Model Interpretability, Analysis and Explanation, Social and Ethical Considerations, Scaling Language Models, Model Quantization, Model Distillation, PEFT.	09
Unit-VI	**Applications of NLC:** Sequence Classification, Token Classification, Summarization, Question-Answering, Coreference Resolution, Code Generation, Low-resource Machine Translation.	07

Self-Study:
The self-study contents will be declared at the commencement of the semester. Around 10% of the questions will be asked from self-study content.

Suggested Readings/ References:

1. J Dan Jurafsky and James H. Martin, Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition, Prentice Hall
2. Jacob Eisenstein, Introduction to Natural Language Processing, The MIT Press
3. Ian Goodfellow, Yoshua Bengio, and Aaron Courville. Deep Learning, The MIT Press
4. Lewis Tunstall, Leandro von Werra, and Thomas Wolf, Natural Language Processing with Transformers, O'Reilly.

Suggested List of Experiments:

Sr. No.	Name of Experiments/Exercises	Hours
1	Introduction to text processing libraries	02
2	Introduction to PyTorch for NLP	02
3	Implement Word2Vec models	04
4	Implement sequence models for (i) Sequence Classification (ii) Named-Entity Recognition (iii) Machine Translation	06
5	Enhance the Seq2Seq model using different attention mechanisms	02
6	Implement transformer model for (i) sequence classification (ii) machine translation	04
7	Fine-tune pre-trained transformer models such as BERT, GPT-2, or T5 on a specific NLP task	02
8	Implement SHAP and LIME for interpretability of NLP models using techniques	02
9	Build a conversational agent using LangChain	04
10	Incorporate RAG/LoRA/RLHF in a conversational agent.	02