NIRMA UNIVERSITY

Institute:	Institute of Technology
Name of Programme:	Integrated B.Tech.(CSE)-MBA
Course Code:	XXXX
Course Title:	Natural Language Computing
Course Type:	Department Elective-III
Year of Introduction:	2024-25

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. infer about major issues and solutions related to natural language computing (BL2)
2. utilize various computational methods to understand language phenomena (BL3)
3. assess the sequence modelling techniques for various use cases (BL5)
4. develop applications with natural language capabilities (BL6)

Unit	Contents	Teaching Hours (Total 45)
Unit-I	**Introduction and Text Classification:** NLP overview, Regular Expressions, Text pre-processing, feature extraction from text, Neural networks for words and characters, Text Mining case study	07
Unit-II	**Language Modelling**: N gram models, Smoothing, Part of speech tagging, Hidden Markov models, Viterbi algorithm, Forward - backward algorithm, EM training, Models for Named Entity Recognition and Part of Speech tagging	18
Unit-III	**Vector Space Models:** Matrix factorization, Word2Vec and Doc2Vec, GloVe, Word - character and sentence embeddings, Topic modelling	06
Unit-IV	**Neural Language Models:** Recurrent Neural Networks and Long Short-term Memory networks, Large Language Models, Generative AI concepts,	07
Unit-V	**Use cases in Sequence Modelling**: Introducing machine translation, Encoder-decoder architecture, Attention mechanism, implementing a conversational chat-bot, Transformers, automatic speech recognition and text to speech synthesis	07

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

Suggested Readings/ References:

1. Manning, Christopher D., and Hinrich Schütze. Foundations of Statistical Natural Language Processing. Cambridge, MA: MIT Press
2. Jurafsky, David, and James H. Martin. Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition. Upper Saddle River, NJ: Prentice-Hall

3. James Allen. Natural Language Understanding. The Benajmins/Cummings Publishing Company Inc.
4. Steven Bird, Ewan Klein, and Edward Loper. Natural Language Processing with Python – Analyzing Text with the Natural Language Toolkit
5. Jacob Perkins, Python Text Processing with NLTK 2.0 Cookbook, Packt Publishing

Suggested List of Experiments:

Sr. No.	Title	Hours
1	Basic Regular Expressions hands on and exploring libraries (NLTK, SpaCy, Gensim) for NLP tasks.	04
2	Implementation of feature representation from text data using word embedding models.	02
3	Implementation of PoS Tagging.	02
4	Implementation of Sentiment Analysis on Twitter dataset.	02
5	Virtual labs on Language models and N-gram detection.	02
6	Implementation of Forward algorithm to address the likelihood problem of hidden Markov models.	04
7	Implementation of Viterbi algorithm to address the decoding problem in hidden Markov models.	02
8	Implementation of basic recurrent neural network for sequence learning task.	04
9	Implementation of machine translation.	04
10	Toy implementation of chatbot / Question Answering system.	04

Suggested Case List: -NA-