Course Title: Natural Language Processing			
Course Code: UDS 308	**No. of credits: 4**	**L-T-P:** 30-15-30	**Learning hours: 60**
			L: Lectures**; T:** Tutorials**; P:** Practicals
Pre-requisite (if any): Data Mining & Analysis (UDS 205), Machine Learning and Applications (UDS 301)			
Department: Natural and Applied Sciences			
Course Coordinator: Dr. Adwitiya Sinha/ Dr Priyanka Singh		**Course Instructor:**	
Contact Details:			
Course Type: Major		**Course Offered in:** Semester-6	

Course Description

This course introduces the core concepts, algorithms, and applications of Natural Language Processing in building intelligent language-based systems. It covers text pre-processing, feature extraction, language modeling, word embeddings, and advanced deep learning architectures such as RNNs, LSTMs, and Transformers. Through hands-on programming using Python and libraries like NLTK, spaCy, Hugging Face Transformers, and Gensim, students will gain practical experience in tasks such as sentiment analysis, text classification, and chatbot development. By the end, students will understand how to integrate linguistic theory with computational models, enabling them to design AI-driven solutions for real-world language understanding and generation.

Course Objectives

- Understand the fundamental concepts of natural language processing and core NLP techniques
- Apply linguistic and deep learning models for text analysis, understanding, and generation
- Design NLP-based applications such as sentiment analysis, translation, and chatbots

Course Content

Module	Topic	L	T	P
1	**Introduction to Natural Language Processing**			
	This module focuses on the fundamentals, importance, and scope of NLP in artificial intelligence. Students will learn how machines process, interpret, and generate natural language through computational models: Introduction to Artificial Intelligence, Definition and scope of NLP, Levels of language analysis (phonology, morphology, syntax, semantics, pragmatics), Phases of Natural Language Processing, NLP applications, Challenges in NLP, Overview of text-based AI systems, Python environment setup for NLP using NLTK and spaCy.	6	2	4
2	**Model Building & Performance Analysis**			
	This module introduces methods for preparing and representing textual data for machine learning and deep learning models. Emphasis is placed on tokenization, normalization, and transforming raw text into numerical features.:	8	3	6

	Importance of Text Pre-processing, tokenization, stop-words removal, stemming, lemmatization, Bag-of-Words, TF-IDF, n-grams, POS tagging, Named Entity Recognition (NER), Word embeddings (Word2Vec, GloVe, FastText), n-gram analysis, document representation techniques			
3	**NLP with Machine Learning and Deep Learning Models**			
	Students will explore classical and neural approaches for text classification and understanding. The module includes implementing traditional ML algorithms and advanced deep learning architectures for NLP tasks. Text classification, Sentiment analysis, Sequence modeling, RNN, LSTM, GRU, Attention mechanism, Transformer architecture, BERT and GPT models, Fine-tuning pre-trained language models, Evaluation metrics, BLEU, METOIR	8	5	10
4	**Applied NLP and Real-World Applications**			
	This module focuses on applying NLP techniques to real-world problems, integrating models into AI-based applications. Students will work on end-to-end projects demonstrating language understanding and generation capabilities. Chatbots and conversational AI, Machine translation, Summarization, Question answering systems, Applications in the filed Customer service chatbots, Market intelligence, Healthcare, Financial sectors, etc. Information retrieval, Ethical considerations in NLP, Bias and fairness, Capstone NLP project using Hugging Face Transformers and Streamlit.	8	5	10
	Total	**30**	**15**	**30**

Practical Sessions	NLP Pre-processing Basics using NLTK and spaCy, including setting environment, text normalization and text cleaning, and linguistic annotation	-	-	4
	Feature engineering including tokenization, bag-of-words, n-gram, and WordCloud, and Word Embeddings using NLTK			6
	Hands-on with text classification using machine learning and deep learning models techniques using Python programming.	-	-	10
	Transformers and fine-tuning BERT using Hugging Face Transformers, Streamlit and Gradio Python Libraries	-	-	10
	Total Practical Sessions	**-**	**-**	**30**

Evaluation Criteria

- Minor Test: Written test [at the end of teaching of modules 1 and 2] -- 20%
- Major Test: Written test [end of semester, full syllabus, modules 1, 2, 3, and 4] -- 40%
- Practical Test: Practical test [including modules 2, 3 and 4] -- 20%
- Project-based learning: Project demonstration and report [at end of module 4] -- 20%

Learning Outcomes

By the end of the course, students will:

- Explain the fundamental concepts in Natural Language Processing, perform text preprocessing, generate and analyze text features such as Bag-of-Words, TF-IDF, n-grams, and linguistic annotations (POS tags and Named Entities). [Module 1 and 2; Minor Test]
- Build, train, and evaluate machine learning and deep learning models for NLP tasks. [Module 2 and 3; Practical Test]
- Design and deploy an end-to-end NLP application integrating linguistic theory with computational models. [Module 1, 2, 3, and 4; Project-based learning and Major Test]

Pedagogical Approach

- This course will provide blended learning approach - Hands-on, Practice-Oriented Learning: Each concept is reinforced through guided coding labs, real datasets, and incremental mini-projects.
- This course is dedicated for incremental learning moving from Model-to-Application Progression. Students move from understanding theory → implementing models → developing complete NLP applications.
- This course will have formative assessments (mini-projects, coding exercises), summative assessments (Mid-term and final project evaluations), and continuous evaluation (online coding platforms for auto-grading assignments).

Reading Resources Recommended (* = suggested readings)

- * James Whitaker, John Chih Liu, and Uday Kamath. (2019). Deep Learning for NLP and Speech Recognition.
- *Jurafsky, D. & Martin, J. H. (2023). Speech and Language Processing (3rd Ed. draft).
- *Bird, S., Klein, E., & Loper, E. (2009). Natural Language Processing with Python.
- Chollet, F. (2021). Deep Learning with Python (2nd Ed.).
- Goldberg, Y. (2017). Neural Network Methods for Natural Language Processing.

Official Documentations of Tools/Libraries

- NLTK - https://www.nltk.org/
- spaCy - https://spacy.io/
- Hugging Face Transformers - https://huggingface.co/transformers
- Gensim - https://radimrehurek.com/gensim/

Student Responsibilities: The students are required:

- To be regular and attentive in theory classes and lab sessions
- To come prepared with readings that would be given in the theory class
- To come prepared in lab with readings from respective theory concepts taught in the class
- To sincerely participate in the day-to-day lab activities
- To incorporate real-word datasets as case-studies in project-based learning
- To submit assignments and project works on or before the deadlines

Course Designed by:

- Dr. Adwitiya Sinha, Associate Professor, Department of Natural and Applied Sciences, TERI School of Advanced Studies, New Delhi
- Dr Priyanka Singh, Assistant Professor, Department of Natural and Applied Sciences, TERI School of Advanced Studies, New Delhi

Course Reviewers:

- Dr. D.K. Lobiyal, Professor, School of Computer and Systems Sciences, Jawaharlal Nehru University, New Delhi
- Dr. Sanjoy Das, Professor and Head, Dept. of Computer Science, Indira Gandhi National Tribal University, Amarkantak