NATURAL LANGUAGE PROCESSING			
Course Code	**22AIM53**	**CIE Marks**	**50**
L:T:P:S	**3:0:0:0**	**SEE Marks**	**50**
Hrs. / Week	**3**	**Total Marks**	**100**
Credits	**03**	**Exam Hours**	**03**

Course outcomes: At the end of the course, the student will be able to:

22AIM53.1	Understand basics of linguistics, probability and statistics associated with NLP.
22AIM53.2	Analyze the semantic of natural language.
22AIM53.3	Design an end-to-end NLP application by integrating preprocessing, feature extraction, and model-building techniques.
22AIM53.4	Evaluate the performance of advanced transformer models (e.g., BERT, GPT-3) in various NLP tasks such as text classification, summarization, and topic modeling.
22AIM53.5	Demonstrate the working of sequence models for text processing.
22AIM53.6	Implement the NLP applications on emerging trends with ethical implications.

Mapping of Course Outcomes to Program Outcomes and Program Specific Outcomes:

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2
22AIM53.1	2	-	-	-	-	-	-	-	-	-	-	-	---	-
22AIM53.2	-	3	-	-	-	-	-	-	-	-	-	2	3	2
22AIM53.3	-	-	3	-	-	-	-	-	-	-	-	2	3	2
22AIM53.4	-	-	3	-		-	-	-	-	-	-	2	3	2
22AIM53.5	-	-	3	-		-	-	-	-	-	-	2	3	2
22AIM53.6	-	-	3	-	3	-	-	2		-	-	2	3	2

MODULE-1	**Natural Language Processing**	22AIM53.1	**8 Hours**

Components - Basics of Linguistics and Probability and Statistics – Words-Tokenization-Morphology: Inflectional Morphology - Derivational Morphology. Finite-State Morphological Parsing - Porter Stemmer.

Case Study	Case studies of NLP applications in various industries.
Text Book	Text Book 1: Ch 2,3,4

MODULE-2	**Semantic Analysis**	**2**2AIM53.2	**8 Hours**

Representing Meaning-Meaning Structure of Language-First Order Predicate Calculus Representing Linguistically Relevant Concepts -Syntax-Driven Semantic Analysis - Semantic Attachments -Syntax-Driven Analyzer. Robust Analysis - Lexemes and Their Senses - Internal Structure - Word Sense Disambiguation -Information Retrieval

Text Book	Text Book 1: 13,14,18

MODULE-3	**WORD REPRESENTATION AND PART OF SPEECH**	22AIM53.2, 22AIM53.3	**8 Hours**

N-grams and Language models –Smoothing- Evaluating Language Model -Text classification- Naïve Bayes classifier –- Vector Semantics – TF-IDF – Word Embeddings: Word2Vec, Glove and Fast Text-Part of Speech – Part of Speech Tagging -Named Entities –Named Entity Tagging-Conditional Random Fields(CRFs).

Text Book	Text Book 1: Ch 4,5,10,17,19

MODULE-4	**Transformer and Topic Models**	22AIM53.4, 22AIM53.5	**8 Hours**

Introduction to transformer architecture-BERT (Bidirectional Encoder Representations from Transformers)-GPT-3 (Generative Pre-trained Transformer 3)-Fine-tuning transformer models for NLP tasks. **Topic Modeling:** Introduction to topic modeling-Latent Dirichlet Allocation (LDA)-Non-Negative Matrix Factorization (NMF).

Text Book	Text Book 1:16,18

MODULE-5	**Applications and Future Directions in NLP**	22AIM53.5, 22AIM53.6	**8 Hours**
Applications and Implementation of NLP: Sentiment Analysis - Text Classification- Text			

Summarization- Named Entity Recognition code- Chatbots and Dialogue systems. **Future Trends in NLP-**Emerging trends and research areas-AI-driven NLP tools and services.

Case Study	Using NLP for Healthcare summaries
Text Book	**Text Book 1: 17-20**

CIE Assessment Pattern (50 Marks – Theory)

RBT Levels		Test	Assessment(s) *	MCQ
		25	**15**	**10**
L1	**Remember**	5		**5**
L2	**Understand**	5	**-**	**5**
L3	**Apply**	10	**5**	
L4	**Analyze**	5	**10**	
L5	**Evaluate**	-	**-**	
L6	**Create**	-	**-**	

*Assessments are to be selected from the assessment list attached to **Appendix A.**.

SEE Assessment Pattern (50 Marks – Theory)

RBT Levels		Exam Marks Distribution (50)
L1	**Remember**	10
L2	**Understand**	10
L3	**Apply**	20
L4	**Analyze**	10
L5	**Evaluate**	-
L6	**Create**	-

Suggested Learning Resources:

Text Books:

1) Daniel Jurafsky and James H. Martin, “Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition (Prentice Hall Series in Artificial Intelligence), 2017. ISBN: 0133252930, 9780133252934

2) Jacob Eisenstein. “Natural Language Processing “, MIT Press, 2019. ISBN: 9780262042840
https://web.stanford.edu/~jurafsky/slp3/(Updated Text book content available link)

Reference Books:

1) Samuel Burns “Natural Language Processing: A Quick Introduction to NLP with Python and NLTK, 2019.

2) Christopher Manning, “Foundations of Statistical Natural Language Processing”, MIT Press, 2009

Web links and Video Lectures (e-Resources):

- https://archive.nptel.ac.in/courses/106/106/106106211/
- https://www.nptelvideos.com/course.php?id=424
- https://www.youtube.com/watch?v=rmVRLeJRkl4

Activity-Based Learning (Suggested Activities in Class)/ Practical Based learning

- Online Class using Jeopardy
- Contents related activities (Activity-based discussions)
  - For active participation of students, instruct the students to read research topics on NLP
  - Class Presentation.