

## NATURAL LANGUAGE PROCESSING (CS566)

### 1. General

1.1 Course Title: Natural Language Processing

1.2 Course Number: CS566

1.2 Contact Hours: 3-0-0 (42 Hours)

1.3 Credits: 9

1.4 Semester Offered: 6<sup>th</sup> semester (Even)

1.5 Prerequisite: Probability, Linear Algebra, Basic Machine Learning, Deep Learning, Python Programming

Evaluation: Participants would be assessed through their involvement in class discussions, quizzes, submission of assignments, mid-term and end-term examinations. The tentative weights for these segments would be as follows:

|                                 |      |
|---------------------------------|------|
| Quizzes/class tests/Assignments | 20 % |
| Mid-Sem                         | 30 % |
| End Sem                         | 50 % |

### 2. Objective:

The objective of this course is to provide a comprehensive understanding of Natural Language Processing (NLP) techniques ranging from classical statistical methods to modern deep learning and transformer-based approaches. The course emphasizes both theoretical foundations and practical implementation for real-world language applications.

### 3. Course Content

Unit-wise distribution of content and number of lectures

| Unit | Topics                            | Sub-topics   | Lectures |
|------|-----------------------------------|--|----------|
| 1    | Introduction to NLP               | History of NLP, NLP pipeline, applications, challenges, linguistic levels, tokenization, normalization, stemming, lemmatization, stop-word removal | 6        |
| 2    | Statistical NLP & Language Models | N-gram language models, smoothing  | 8        |

|   |  |  |   |
|---|--|--|---|
|   |  | techniques, entropy, perplexity, word frequency analysis, , TF-IDF, feature extraction   |   |
| 3 | Text Representation & Classical Models | Bag-of-Words, word embeddings (Word2Vec, GloVe), similarity measures, text classification, sentiment analysis using Naive Bayes, Logistic Regression and SVM | 8 |
| 4 | Sequence Modeling & Deep Learning      | Neural language models, RNN, LSTM, GRU, sequence-to-sequence models, attention mechanism, applications in machine translation and summarization              | 8 |
| 5 | Transformer-based NLP                  | Self-attention, Transformer architecture, BERT, GPT, RoBERTa, pretraining and fine-tuning, prompt-based learning, transfer learning                          | 8 |
| 6 | Advanced Topics & Applications         | Named Entity Recognition, POS tagging, dependency parsing, question answering, dialogue systems, multilingual NLP, ethical issues, bias, and fairness in NLP | 4 |

Total Lectures: 42

#### **4. Readings**

##### 4.1 Textbooks:

1. Jurafsky, D. and Martin, J. H., Speech and Language Processing, Pearson.
2. Eisenstein, J., Introduction to Natural Language Processing, MIT Press.

##### 4.2 Reference Books:

1. Goldberg, Y., Neural Network Methods for Natural Language Processing, Morgan & Claypool.
2. Goodfellow, I., Bengio, Y., Courville, A., Deep Learning, MIT Press.

#### **5. Outcome of the Course**

After successful completion of this course, students will be able to:

- Understand linguistic and statistical foundations of NLP
- Implement classical and deep learning-based NLP models
- Apply transformer-based models for real-world NLP tasks
- Critically analyze ethical and societal issues related to NLP systems