

|   |   |  |                             |                       |                               |                 |   |                       |  |     |  |  |  |   |   |   |   |   |   |   |
|---|---|--|-----------------------------|-----------------------|-------------------------------|-----------------|---|-----------------------|--|-----|--|--|--|---|---|---|---|---|---|---|
| Course Code   | 21CSE356T   | Course Name  | NATURAL LANGUAGE PROCESSING |                       |                               | Course Category | E | PROFESSIONAL ELECTIVE |  |     |  |  |  | L | T | P | C |   |   |   |
|   |   |  |                             |                       |                               |                 |   |                       |  |     |  |  |  |   |   |   | 2 | 1 | 0 | 3 |
| Pre-requisite Courses   |   | Nil  |                             | Co- requisite Courses |                               | Nil             |   | Progressive Courses   |  | Nil |  |  |  |   |   |   |   |   |   |   |
| Course Offering Department  |   | School of Computing                                  |                             |                       | Data Book / Codes / Standards |                 |   | Nil                   |  |     |  |  |  |   |   |   |   |   |   |   |
| Course Learning Rationale (CLR):  |   | The purpose of learning this course is to:           |                             |                       |                               |                 |   |                       |  |     |  |  |  |   |   |   |   |   |   |   |
| CLR-1:  | Understand the fundamentals behind the Language processing and perform word level analysis.                       |  |                             |                       |                               |                 |   |                       |  |     |  |  |  |   |   |   |   |   |   |   |
| CLR-2:  | Understand the syntactic processing and probabilistic context-free grammars.                                      |  |                             |                       |                               |                 |   |                       |  |     |  |  |  |   |   |   |   |   |   |   |
| CLR-3:  | Conceive the basics of the knowledge representation, inference, and discourse analysis.                           |  |                             |                       |                               |                 |   |                       |  |     |  |  |  |   |   |   |   |   |   |   |
| CLR-4:  | Recognize the significance of transformer-based models.   |  |                             |                       |                               |                 |   |                       |  |     |  |  |  |   |   |   |   |   |   |   |
| CLR-5:  | Understand the natural language processing applications and to learn how to apply basic algorithms in this field. |  |                             |                       |                               |                 |   |                       |  |     |  |  |  |   |   |   |   |   |   |   |
| Course Outcomes (CO):   |   | At the end of this course, learners will be able to: |                             |                       |                               |                 |   |                       |  |     |  |  |  |   |   |   |   |   |   |   |
| CO-1:   | Exhibit knowledge on text preprocessing techniques and perform word level analysis.                               |  |                             |                       |                               |                 |   |                       |  |     |  |  |  |   |   |   |   |   |   |   |
| CO-2:   | Illustrate approaches to syntax analysis including probabilistic context-free grammars                            |  |                             |                       |                               |                 |   |                       |  |     |  |  |  |   |   |   |   |   |   |   |
| CO-3:   | Apply approaches to semantics and discourse analysis in NLP.  |  |                             |                       |                               |                 |   |                       |  |     |  |  |  |   |   |   |   |   |   |   |
| CO-4:   | Develop models using transfer learning approaches.  |  |                             |                       |                               |                 |   |                       |  |     |  |  |  |   |   |   |   |   |   |   |
| CO-5:   | Implement applications that use Natural Language Processing approaches.   |  |                             |                       |                               |                 |   |                       |  |     |  |  |  |   |   |   |   |   |   |   |
| Unit-1 - Overview and Word Level Analysis9 Hour   |   |  |                             |                       |                               |                 |   |                       |  |     |  |  |  |   |   |   |   |   |   |   |
| Introduction to Natural Language Processing, Applications of NLP, Levels of NLP, Regular Expressions, Morphological Analysis, Tokenization, Stemming, Lemmatization, Feature extraction: Term Frequency (TF), Inverse Document Frequency (IDF), Modeling using TF-IDF, Parts of Speech Tagging, Named Entity Recognition, N-grams, Smoothing. |   |  |                             |                       |                               |                 |   |                       |  |     |  |  |  |   |   |   |   |   |   |   |
| Unit-2 - Syntax Analysis9 Hour  |   |  |                             |                       |                               |                 |   |                       |  |     |  |  |  |   |   |   |   |   |   |   |
| Context Free Grammars, Grammar Rules for English, Top-Down Parsing, Bottom-Up Parsing, Ambiguity, CKY Parsing, Dependency Parsing, Earley Parsing - Probabilistic Context-Free Grammars   |   |  |                             |                       |                               |                 |   |                       |  |     |  |  |  |   |   |   |   |   |   |   |
| Unit-3 - Semantic and Discourse Analysis9 Hour  |   |  |                             |                       |                               |                 |   |                       |  |     |  |  |  |   |   |   |   |   |   |   |
| Representing Meaning, Lexical Semantics, Word Senses, Relation between Senses, Word Sense Disambiguation, Word Embeddings, Word2Vec, CBOW, Skip-gram and GloVe, Discourse Segmentation, Text Coherence, Discourse Structure, Reference Resolution, Pronominal Anaphora Resolution, Coreference Resolution                                     |   |  |                             |                       |                               |                 |   |                       |  |     |  |  |  |   |   |   |   |   |   |   |
| Unit-4 - Language Models9 Hour  |   |  |                             |                       |                               |                 |   |                       |  |     |  |  |  |   |   |   |   |   |   |   |
| Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM), Attention mechanism, Transformer Based Models, Self-attention, multi-headed attention, BERT, RoBERTa, Fine Tuning for downstream tasks, Text classification and Text generation.  |   |  |                             |                       |                               |                 |   |                       |  |     |  |  |  |   |   |   |   |   |   |   |
| Unit-5 - NLP Applications9 Hour   |   |  |                             |                       |                               |                 |   |                       |  |     |  |  |  |   |   |   |   |   |   |   |
| Introduction to Chatbot Applications, Retrieval based- Conversation based, Information Extraction and its approaches, Information Retrieval, Semantic Search and Evaluation, Question Answering, Summarization, Extractive Vs Abstractive Summarization, Machine Translation.   |   |  |                             |                       |                               |                 |   |                       |  |     |  |  |  |   |   |   |   |   |   |   |

|                           |  |  |
|---------------------------|--|--|
| <b>Learning Resources</b> | 1. Daniel Jurafsky and James H Martin, "Speech and Language Processing: An introduction to Natural Language Processing, Computational Linguistics and Speech Recognition", Prentice Hall, 2nd Edition, 2018. | 4. Rothman, Denis. Transformers for Natural Language Processing: Build innovative deep neural network architectures for NLP with Python, PyTorch, TensorFlow, BERT, RoBERTa, and more. Packt Publishing Ltd, 2021. |
|                           | 2. C.Manning and H.Schutze, —Foundations of Statistical Natural Language Processingll, MIT Press. Cambridge, MA, 1999  | 5. <a href="http://mccormickml.com/2106/04/19/word2vec-tutorial-the-skip-gram-model/">http://mccormickml.com/2106/04/19/word2vec-tutorial-the-skip-gram-model/</a>   |
|                           | 3. JamesAllen, Bejamin/cummings, — NaturalLanguageUnderstandingll, 2ndedition, 1995  | 6. <a href="https://nlp.stanford.edu/pubs/glove.pdf">https://nlp.stanford.edu/pubs/glove.pdf</a>   |

| Learning Assessment |                              |  |          |                                      |          |   |          |
|---------------------|------------------------------|--|----------|--------------------------------------|----------|---|----------|
|                     | Bloom's<br>Level of Thinking | Continuous Learning Assessment (CLA)             |          |                                      |          | Summative<br>Final Examination<br>(40% weightage) |          |
|                     |                              | Formative<br>CLA-1 Average of unit test<br>(50%) |          | Life-Long Learning<br>CLA-2<br>(10%) |          |   |          |
|                     |                              | Theory   | Practice | Theory                               | Practice | Theory  | Practice |
| Level 1             | Remember                     | 15%  | -        | 15%                                  | -        | 15%   | -        |
| Level 2             | Understand                   | 25%  | -        | 25%                                  | -        | 25%   | -        |
| Level 3             | Apply                        | 30%  | -        | 30%                                  | -        | 30%   | -        |
| Level 4             | Analyze                      | 30%  | -        | 30%                                  | -        | 30%   | -        |
| Level 5             | Evaluate                     | -  | -        | -                                    | -        | -   | -        |
| Level 6             | Create                       | -  | -        | -                                    | -        | -   | -        |
|                     | Total                        | 100 %  |          | 100 %                                |          | 100 %   |          |

| <b>Course Designers</b>  |  |                                |
|--|--|--------------------------------|
| Experts from Industry  | Experts from Higher Technical Institutions                                       | Internal Experts               |
| 1. Dr. J.Balaji, Associate Manager, Allstate Solutions Pvt Ltd, <a href="mailto:jagank.balaji@gmail.com">jagank.balaji@gmail.com</a> | 1. Dr. Vani. V, Assistant Professor, National Institute of Technology Puducherry | 1. Dr. R. Anita, SRMIST.       |
|  |  | 2. Dr.Subalalitha C.N , SRMIST |
|  |  | 3. Ms.Viji D , SRMIST          |