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:*	Program Outcomes (PO)												Program Specific Outcomes		
		1	2	3	4	5	6	7	8	9	10	11	12			
		Engineering Knowledge	Problem Analysis	Design/development of solutions	Conduct investigations of complex problems	Modern Tool Usage	The engineer and society	Environment & Sustainability	Ethics	Individual & Team Work	Communication	Project Mgt. & Finance	Life Long Learning	PSO-1	PSO-2	PSO-3
CO-1:	*exhibit knowledge on text preprocessing techniques and perform word level analysis*	*3*	*3*	*2*	-	-	-	-	-	-	-	-	-	*2*	-	-
CO-2:	*illustrate approaches to syntax analysis including probabilistic context-free grammars*	*3*	*3*	*2*	-	-	-	-	-	-	-	-	-	*2*	-	-
CO-3:	*apply approaches to semantics and discourse analysis in NLP*	*3*	*3*	*2*	-	-	-	-	-	-	-	-	-	*2*	-	-
CO-4:	*develop models using transfer learning approaches*	*3*	-	-	*3*	*3*	-	-	-	-	-	-	-	*2*	-	-
CO-5:	*implement applications that use Natural Language Processing approaches*	-	-	*2*	*3*	*3*	-	-	-	-	-	-	-	-	-	-

Unit-1 - Overview and Word Level Analysis	***9 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 Analysis	***9 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 Analysis	***9 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 Models	***9 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 Applications	***9 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.	

Learning Resources	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.* 2. *C.Manning and H.Schutze, ―Foundations of Statistical Natural Language Processing‖, MIT Press. Cambridge, MA,1999* 3. *JamesAllen, Bejamin/cummings, ―NaturalLanguageUnderstanding‖,2ndedition,1995*	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.* 5. *http://mccormickml.com/2106/04/19/word2vec- tutorial-the-skip-gram-model/* 6. *https://nlp.stanford.edu/pubs/glove.pdf*

Learning Assessment

	Bloom's Level of Thinking	*Continuous Learning Assessment (CLA)*				*Summative Final Examination (40% weightage)*	
		Formative CLA-1 Average of unit test (50%)		*Life-Long Learning CLA-2 (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 %*	

Course Designers

Experts from Industry	**Experts from Higher Technical Institutions**	**Internal Experts**
1. *Dr. J.Balaji, Associate Manager,Allstate Solutions Pvt Ltd, jagank.balaji@gmail.com*	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*