FACULTY OF ENGINEERING & TECHNOLOGY

Effective from Academic Batch: 2022-23

Programme: Bachelor of Technology (Artificial Intelligence(AI) and Data Science)

Semester: VII

Course Code: 202047809

Course Title: Natural Language Processing

Course Group: Professional Elective Course - IV

Course Objectives: This course deals with processing text data using probabilistic model with n-grams and preprocessing techniques. Concepts of text classification, text summarization, semantic analysis drives the subject to machine translation. This course will boost student knowledge to a level that can help to develop NLP applications like surveys, chatbots QA system etc. which is required in ecommerce platforms.

Teaching & Examination Scheme:

Contact hours per week			Course Credits	Examination Marks (Maximum / Passing)				
				Theory		J/V/P*		Total
Lecture	Tutorial	Practical		Internal	External	Internal	External	
3	0	2	4	50 / 18	50 / 17	25 /09	25 /09	150 / 53

* **J**: Jury; **V**: Viva; **P**: Practical

Detailed Syllabus:

Sr.	Contents	Hours
1	**Introduction to NLP:** What is NLP? Challenges of NLP, History of NLP, Advantages of NLP, Disadvantages of NLP, Components of NLP, Applications of NLP, how to build an NLP pipeline? Phases of NLP, NLP APIs, NLP Libraries.	04
2	**Natural Language Processing Models and Algorithm:** Unigram Language Model, Bigram, Trigram, N-gram, Advanced smoothing for language modelling, Applications of Language Modeling, Natural Language Generation, Parts of Speech Tagging, Morphology, Named Entity Recognition	06
3	**Text Processing:** Continuous Bag-Of-Words, embedding representations for words Lexical Semantics, Word Sense Disambiguation, Knowledge-Based and Supervised Word Sense Disambiguation, Tokenization, Cleaning, Tokenizing, Removing Special Characters, Expanding Contractions, Removing Stop words, Correcting Words, Stemming, Lemmatization, Understanding Text structure.	10

4	**Text Analysis, Summarization and Extraction:** Text Classification, Text Summarization, Information Extraction, Named Entity Recognition, Question Answering in Multilingual Setting; NLP in Information Retrieval, Cross-Lingual IR	**06**
5	**Applications of NLP:** Unsupervised Learning on Text Clustering by Document Similarity – Distance Metrics, Partitive Clustering, Hierarchical Clustering: Analyzing Document Similarity, Document Clustering, Speech recognition.	**06**
6	**Machine Translation:** Need of MT, Problems of Machine Translation, MT Approaches, Direct Machine Translations, Rule-Based Machine Translation, Knowledge-Based MT System, Statistical Machine Translation (SMT)	**08**
	Total	**40**

List of Practicals / Tutorials:

1	Introduction to python libraries for feature extraction and NLP.
2	Apply preprocessing steps on the selected Dataset. Support link for data set: https://www.kaggle.com/learn/natural-language-processing
3	Tokenization: Split the text sentence/paragraph/Data set and generate Tokens.
4	Implement a suitable stemming algorithm based on chosen data set
5	POS tagging part 1: Perform POS tagging annotation on input text.
6	POS tagging part 2: Analyze the result of POS Tagging.
7	Generate N-grams of the text.
8	Convert text into TF IDF vectors.
9	Perform text classification.
10	Implement text similarity technique.
11	Case Study – Study and prepare a report of Machine translation of Google.
12	Case Study - Identify the sentiment of tweets
Note*: All practical can be performed on real-time data set available online: https://www.kaggle.com/learn/natural-language-processing	

Reference Books:

1	Speech and Language Processing - An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition Second Edition by Daniel Jurafsky and James H. Martin, Pearson Education India
2	Foundations of Statistical Natural Language Processing, Chris Manning and Hinrich Schütze, MIT Press
3	Text analytics with python, Dipanjan Sarkar, Apress
4	Computational Nonlinear Morphology: With Emphasis on Semitic Languages, Kiraz, George Anton; Cambridge University Press
5	Oxford Handbook of Computational Linguistics.
6	Natural Language Understanding, James Allen. The Benjamin/Cummings Publishing Company Inc.
7	Natural Language Processing with Python – Analyzing Text with the Natural Language Toolkit, Steven Bird, Ewan Klein, and Edward Loper.

Supplementary learning Material:	
1	https://nptel.ac.in/
2	https://nptel.ac.in/courses/106/105/106105158/
3	https://www.kaggle.com/learn/natural-language-processing
4	https://www.javatpoint.com/nlp

Pedagogy:
- Direct classroom teaching
- Audio Visual presentations/demonstrations
- Assignments/Quiz
- Continuous assessment
- Interactive methods
- Seminar/Poster Presentation
- Industrial/ Field visits
- Course Projects

Suggested Specification table with Marks (Theory) (Revised Bloom's Taxonomy):

Distribution of Theory Marks in %						**R**: Remembering; **U**: Understanding; **A**: Applying;
R	**U**	**A**	**N**	**E**	**C**	**N**: Analyzing; **E**: Evaluating; **C**: Creating
20%	30%	20%	10%	10%	10%	

Note: This specification table shall be treated as a general guideline for students and teachers. The actual distribution of marks in the question paper may vary slightly from above table.

Course Outcomes (CO):

Sr.	Course Outcome Statements	%weightage
CO-1	Understand the basic concepts and requirements of NLP.	**20**
CO-2	Understand and apply various NLP Models and Algorithms.	**30**
CO-3	Design and develop basic applications for text or information extraction.	**30**
CO-4	Apply Machine translation techniques for language translation.	**20**

Curriculum Revision:	
Version:	2.0
Drafted on (Month-Year):	June-2022
Last Reviewed on (Month-Year):	-
Next Review on (Month-Year):	June-2026