Shri Vile Parle Kelavani Mandal's

Dwarkadas J. Sanghvi College of Engineering

(Autonomous College Affiliated to the University of Mumbai)

Scheme and detailed syllabus of DJS23 Honors Program in Data Analytics

With effect from the Academic Year: 2025-2026

Shri Vile Parle Kelavani Mandal's

DWARKADAS J. SANGHVI COLLEGE OF ENGINEERING

(Autonomous College Affiliated to the University of Mumbai)

NAAC Accredited with "A" Grade (CGPA: 3.18)

Department of Computer Science and Engineering (IoT and Cyber Security with Block Chain Technology)

Proposed Scheme for Honors Degree Program in Data Analytics : Semester: V (Autonomous)

(Academic Year 2025-2026)

Sr. No.	Course Code	Course	Teaching Scheme			Continuous Assessment (A)						Semester End Examination (B)						Aggregate (A+B)	Credits
			Theory (hrs.)	Practical (hrs.)	Tutorial (hrs.)	Term Test 1 (TT1) - a	Term Test 2 (TT2) - b	Assg/CP/ GD/Presn tation/Qu iz) - c	Total (a+b+c)	Term work	CA Total	Duration	Theory	Oral	Pract	Oral & Pract	SEE Total		
SEM III																			
1	DJS23BCH1301	Fundamentals of Data Mining	3	--	--	15	15	10	40	--	40	2	60	--	--	--	60	100	3
SEM IV																			
2	DJS23BCH1401	Data Analytics and Visualization	3	--	--	15	15	10	40	--	40	2	60	--	--	--	60	100	3
	DJS23BLH1401	Data Analytics and Visualization Laboratory	--	2	--	--	--	--	--	25	25	2	--	25	--	--	25	50	1
SEM V																			
3	DJS23BCH1501	Natural Language Processing	3	--	--	15	15	10	40	--	40	2	60	--	--	--	60	100	3
	DJS23BLH1501	Natural Language Processing Laboratory	--	2	--	--	--	--	--	25	25	2	--	25	--	--	25	50	1
SEM VI																			
4	DJS23BCH1601	Time Series and Forecasting Analytics	3	--	--	15	15	10	40	--	40	2	60	--	--	--	60	100	3
	DJS23BLH1601	Time Series and Forecasting Analytics Laboratory	--	2	--	--	--	--	--	25	25	2	--	25	--	--	25	50	1
SEM VIII																			
5	DJS23BCH1801	Optimization for Decision Analytics	3	--	--	15	15	10	40	--	40	2	60	--	--	--	60	100	3
		Total	**15**	**6**	**0**	**75**	**75**	**50**	**200**	**75**	**275**	**16**	**300**	**75**	**0**	**0**	**375**	**650**	**18**

Prepared by | Checked by | Head of Department | Vice Principal | Principal

Program: B.Tech in Computer Science and Engineering(IoT and Cybersecurity with Block chain Technology) With Honors in Data Analytics	T.Y.B.Tech	Semester : V
Course : Natural Language Processing	Course Code: DJS23BCH1501	
Course: Natural Language Processing Laboratory	Course Code: DJS23BLH1501	

Teaching Scheme (Hours / week)				Evaluation Scheme								
				Continuous Assessment Marks (A)				Semester End Examination Marks (B)			Total marks (A+B)
Lectures	Practical	Tutorial	Total Credits	Term Test 1	Term Test 2	Assignment	Total	Theory			
				15	15	10	40	60			100
				Term Work				Laboratory Examination			50
3	2	--	4	Laboratory Work	Tutorial / Mini project /presentation / Journal/ Practical	Total Term work		Oral	Practical	Oral & Practical	
				15	10	25		25	–	–	

Pre-requisite:

1. Data Mining
2. Python Programming

Course Objectives: The objectives of the course are:

1. To understand natural language processing and to learn how to apply basic algorithms in this field.
2. To familiarize with the basic concepts and algorithmic description of the main language levels: morphology, syntax, semantics, and pragmatics.
3. To implement various language Models.
4. To design and implement applications based on natural language processing.

Course Outcomes: On completion of the course, learners will be able to:

1. Demonstrate a broad understanding of the field of natural language processing.
2. Comprehend and apply fundamental concepts and algorithmic techniques associated with key language levels.
3. Study different POS tagging techniques, including rule-based and stochastic approaches.
4. Analyze fundamental concepts of lexical semantics and their role in Natural Language Processing.
5. Explore syntactic and semantic constraints on co-reference resolution.
6. Apply NLP techniques to design real world NLP applications.

Detailed Syllabus:		
Unit	**Description**	**Duration**
1	**Introduction:** History of NLP, Evolution of NLP: From Rule-Based Systems to LLMs, Definition and Examples: GPT, BERT, T5, LLaMA, PaLM, Differences between traditional models vs LLMs, Generic NLP system, levels of NLP, Knowledge in language processing, Ambiguity in Natural language, stages in NLP, challenges of NLP, Applications of NLP.	07
2	**Word Level Analysis:** Morphology analysis –survey of English Morphology, Inflectional morphology & Derivational morphology, Lemmatization, Regular expression, finite automata, finite state transducers (FST), Morphological parsing with FST, Lexicon free FST Porter stemmer. N –Grams- N-gram language model, N-gram for spelling correction.	06
3	**Syntax analysis:** Part-Of-Speech tagging (POS)- Tag set for English (Penn Treebank) , Rule based POS tagging, Stochastic POS tagging, Issues –Multiple tags & words, Unknown words. Introduction to CFG, Sequence labeling: Hidden Markov Model (HMM), Maximum Entropy, and Conditional Random Field (CRF).	07
4	**Semantic Analysis:** Lexical Semantics, Attachment for fragment of English- sentences, noun phrases, Verb phrases, prepositional phrases, Relations among lexemes & their senses –Homonymy, Polysemy, Synonymy, Hyponymy, WordNet, Robust Word Sense Disambiguation (WSD), Dictionary based approach	07
5	**Pragmatics:**	06

	Discourse –reference resolution, reference phenomenon, syntactic & semantic constraints on co reference	
6	**Applications:** Machine translation, Information retrieval, Question answers system, categorization, summarization, sentiment analysis, Named Entity Recognition.	06
	Total	**39**

List of Laboratory Experiments:	
Sr. No.	**Suggested Experiments**
1	a. To implement tokenization by word and sentence using nltk. b. To eliminate stopwords using nltk c. To perform stemming using nltk
2	a. To perform Parts of Speech tagging using nltk. b. To implement lemmatization using nltk.
3	a. To implement chunking using nltk. b To implement Named Entity Recognition using nltk.
4	a. To find Term Frequency and Inverse Document Frequency (TF-IDF). b. To implement CYK parsing (CockeYounger-Kasami Parsing) or Chart Parsing.
5	a. To find all unigrams, bigrams and trigrams present in the given corpus. b. To find the probability of the given 31 statement "This is my cat" by taking an example corpus into consideration.
6	Use the Stanford named Entity recognizer to extract entities from the documents. Use it programmatically and output for each document which named entities it contains and of which type.
7	Choose any corpus available on the internet freely. For the corpus, for each document, count how many times each stop word occurs and find out which are the most frequently occurring stop words. Further, calculate the term frequency and inverse document frequency as The motivation behind this is basically to find out how important a document is to a given query. For e.g.: If the query is say: "The brown crow". "The" is less important. "Brown" and "crow" are relatively more important. Since "the" is a more common word, its tf will be high. Hence we multiply it by idf, by knowing how common it is to reduce its weight.

8	To implement sentiment analysis using NLP.
9	To implement Spam Filter using NLP.
10	To detect Fake News using NLP.

Any other experiment may be included, which would help the learner to understand the topic/concept.

Books Recommended:
Text Books:

1. Daniel Jurafsky, James H. Martin , Speech and Language Processing‖ Second Edition, Prentice Hall, 2008.
2. Christopher D.Manning and Hinrich Schutze, Foundations of Statistical Natural Language Processing —, MIT Press, 1999.
3. Brojo Kishore Mishra, Raghvendra Kumar, Natural Language Processing in Artificial Intelligence, Apple Academic Press, 2022.

Reference Books:

1. Siddiqui and Tiwary U.S., Natural Language Processing and Information Retrieval, Oxford University Press, 2008.

2. Daniel M Bikel and Imed Zitouni , Multilingual natural language processing applications‖ Pearson, 2013.

3. Alexander Clark (Editor), Chris Fox (Editor), Shalom Lappin (Editor) , The Handbook of Computational Linguistics and Natural Language Processing , ISBN: 978-1-118-

4. Steven Bird, Ewan Klein, Natural Language Processing with Python, O'Reilly

5. Niel J le Roux, Sugnet Lubbe, A step by step tutorial : An introduction into R application and programming

Web Resources:

1. Natural language toolkit- https://www.nltk.org/
2. Natural language Processing: https://www.deeplearning.ai/resources/natural-language-processing/

Online Courses: NPTEL / Swayam

1. Natural Language Processing by Prof. Pawan Goyal, IIT Kharagpur,
https://onlinecourses.nptel.ac.in/noc21_cs102/preview
2. Applied Natural Language Processing, By Prof. Ramaseshan R, CMI,
https://onlinecourses.nptel.ac.in/noc20_cs87/preview

Evaluation Scheme:

Continuous Assessment (A)

Theory:

1. Term Test 1 (based on 40 % syllabus) of 15 marks for the duration of 45 min.
2. Term Test 2 (on next 40 % syllabus) of 15 marks for the duration of 45 min.
3. Assignment / course project / group discussion /presentation / quiz/ case study for 10 marks.

Laboratory: (Term work)

Term work shall consist of minimum 8 experiments.
The distribution of marks for term work shall be as follows:
i. Laboratory work (Performance of Experiments): 15 Marks
ii. Journal documentation (Write-up and/or Assignments): 10 marks

The final certification and acceptance of term work will be subject to satisfactory performance of laboratory work, and upon fulfilling minimum passing criteria in the term work.

Semester End Examination (B):

Theory:

1. Question paper will be based on the entire syllabus summing up to 60 marks.
2. Total duration allotted for writing the paper is 2 hrs.

Laboratory:

Oral examination will be based on the entire syllabus including, the practical's performed during laboratory sessions.

Prepared by | Checked by | Head of the Department | Vice Principal | Principal