Program	ne BS Computational Statistics and Data Analytics	Course Code	CSTA- 408	Credit Hours	3	
Course Title Natural Language Processing						
Course Introduction						
synthesis o	guage Processing (NLP) is the app f natural language and speech. g experience in Python					
	Learr	ing Outcomes				
 Learn a Master Implem 	of this course, students will be a bout basic NLP problems, tasks basic programming tools for NL ent a simple NLP system. a NLP problem and find a suitab	and methods. P.				
	Course Content		As	ssignments/Readi	ings	
Week 1	Unit – II Understanding the distinction between natural					
Ianguages and formal languages. Unit – III Regular Expressions and Finite State Automata: Basics of regular expressions for pattern matching and text processing. Unit – IV Overview of finite state automata and their applications in language processing.						
Week 3	Unit – V N-grams and Language Models: Explanation of N-grams as consecutive sequences of N items (usually words) in text. Unit – VI Introduction to language models for predicting the likelihood of word sequences.					
Week 4	Unit – VII An Introduction to Programming in Python: Overview of Python programming language, its syntax, and basic concepts. Unit – VIII Setting up Python development environment and writing simple Python programs.					
Week 5	Unit – IX					

	Detailed discussion on Python data types and data	
	structures.	
	Unit - X	
	Hands-on exercises on working with variables,	
	numbers, strings, arrays, and dictionaries in Python.	
	Unit – XI	
Week 6	Conditionals and Iteration in Python:	
	Understanding control flow structures such as if	
	statements, loops, and iteration in Python.	
	Unit – XII	
	Writing Python programs with conditional	
	statements and loops.	
	Unit – XIII	
	The NLTK (Natural Language Toolkit), with	
	Demonstrations:	
Week 7	Overview of the Natural Language Toolkit (NLTK)	
vv cen v	library for natural language processing in Python.	
	Unit – XIV	
	Demonstrations and hands-on exercises on using	
	NLTK for text processing tasks.	
	Unit – XV	
	Part of Speech Tagging:	
Week 8	Introduction to part of speech tagging and its	
WEEK 0	importance in natural language understanding.	
	Unit – XVI	
	Using NLTK for part of speech tagging tasks.	
	Unit – XVII	
	Syntax Parsing:	
Week 9	Basics of syntax parsing for analyzing the	
Week >	grammatical structure of sentences.	
	Unit – XVIII	
	Overview of parsing techniques and algorithms.	
	Unit – XIX	
	Keyword and Phrase Extraction:	
	Techniques for extracting important keywords and	
Week 10	phrases from text documents.	
	Unit – XX	
	Hands-on exercises on implementing keyword and	
	phrase extraction algorithms.	
	Unit – XXI	
	Vector Space Model and Dimensionality Reduction:	
	Introduction to the vector space model for	
Week 11	representing text documents as numerical vectors.	
WEEK II	Unit – XXII	
	Techniques for dimensionality reduction in text data	
	using methods like Singular Value Decomposition	
	(SVD) and Principal Component Analysis (PCA).	
1		
	Unit – XXIII	
Week 12	Unit – XXIII Topic Modeling: Understanding topic modeling algorithms such as	

	Letert Divisibility Allegetics (LDA) (1'			
	Latent Dirichlet Allocation (LDA) for discovering			
	latent topics in text corpora.			
	Unit – XXIV			
	Implementing topic modeling using Python libraries			
	like Gensim.			
	Unit – XXV			
	Distributional Models:			
	Introduction to distributional models for			
	representing word meanings based on their			
Week 13	distributional properties in a corpus.			
	Unit – XXVI			
	Hands-on exercises on building distributional word			
	embeddings using techniques like Word2Vec and			
	GloVe.			
	Unit – XXVII			
	Text Classification:			
	Overview of text classification techniques using			
Week 14	supervised learning algorithms.			
	Unit – XXVIII			
	Using machine learning classifiers to classify text			
	documents into predefined categories.			
	Unit – XXIX			
	Sequence Labeling:			
	Explanation of sequence labeling tasks such as			
XX7.1.1 7	named entity recognition (NER) and part of speech			
Week 15	tagging.			
	Unit – XXX			
	Hands-on exercises on implementing sequence			
	labeling algorithms using machine learning and deep			
	learning approaches.			
	Unit – XXXI Deer Learning for NLD: DOS Teacing:			
	Deep Learning for NLP: POS Tagging:			
	Introduction to deep learning techniques for natural language processing, with a focus on part of speech			
Week 16				
Week 10	tagging. Unit – XXXII			
	Building and training deep learning models for POS tagging using frameworks like TensorFlow or			
	PyTorch.			
Textbooks and Reading Material				
Text Book				
	S., Klein, E., & Loper, E. (2009). Natural languag	e processing with Puthon.		
	ng text with the natural language toolkit. "O'Reilly Med			
unui y2,1	ng iom with the hundral uniguage tootkit. O Kenty Med	iu, 1110		

Suggested Readings

1. Jurafsky, D., & Martin, J. H. (1999). Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. Prentice Hall.

Teaching Learning Strategies

Class Lecture method, which includes seminars, discussions, assignments and projects. (Audio-visual tools are used where necessary)

Assignments: Types and Number with Calendar

According to the choice of respective teacher.

	Assessment				
Sr. No.	Elements	Weightage	Details		
1.	Midterm Assessment	35%	It takes place at the mid-point of the semester.		
2.	Formative Assessment	25%	It is continuous assessment. It includes: Classroom participation, attendance, assignments, and presentations, homework, attitude and behavior, hands-on-activities, short tests, quizzes etc.		
3.	Final Assessment	40%	It takes place at the end of the semester. It is mostly in the form of a test, but owing to the nature of the course the teacher may assess their students based on term paper, research proposal development, field work and report writing etc.		

Programme	BS Computational Statistics and Data Analytics	Course Code	CSTA- 408	Credit Hours	3	
Course Title	Natural Language Processin	Natural Language Processing				
	Course	Introduction				
synthesis of	Natural Language Processing (NLP) is the application of computational techniques to the analysis and synthesis of natural language and speech. This course is an introduction to NLP with prior programming experience in Python					
Learning Outcomes						
•	this course, students will be ab					
	out basic NLP problems, tasks a					
	sic programming tools for NLP	? .				
1	t a simple NLP system.					
8. Define a NLP problem and find a suitable solution to it.						
Course Content Assignments/Readings					ings	
Week 1 Unit – I						

[Natural Language and Formal Language	
	Natural Language and Formal Language:	
	Introduction to natural language and formal	
	language theory. Unit – II	
	Understanding the distinction between natural	
	languages and formal languages.	
	Regular Expressions and Finite State Automata:	
	Basics of regular expressions for pattern matching	
Week 2	and text processing.	
	Unit – IV	
	Overview of finite state automata and their	
	applications in language processing.	
	Unit – V	
	N-grams and Language Models:	
	Explanation of N-grams as consecutive sequences of	
Week 3	N items (usually words) in text.	
	Unit – VI	
	Introduction to language models for predicting the	
	likelihood of word sequences.	
	Unit – VII	
	An Introduction to Programming in Python:	
	Overview of Python programming language, its	
Week 4	syntax, and basic concepts.	
	Unit – VIII	
	Setting up Python development environment and	
	writing simple Python programs.	
	Unit – IX	
	Variables, Numbers, Strings, Arrays, Dictionaries in	
	Python:	
	Detailed discussion on Python data types and data	
Week 5	structures.	
	Unit – X	
	Hands-on exercises on working with variables,	
	numbers, strings, arrays, and dictionaries in Python.	
	Unit – XI	
	Conditionals and Iteration in Python:	
	Understanding control flow structures such as if	
Week 6	statements, loops, and iteration in Python.	
	Unit – XII	
	Writing Python programs with conditional	
	statements and loops.	
	Unit – XIII	
	The NLTK (Natural Language Toolkit), with	
	Demonstrations:	
	Overview of the Natural Language Toolkit (NLTK)	
Week 7	library for natural language processing in Python.	
	Unit – XIV	
	Demonstrations and hands-on exercises on using	
	NLTK for text processing tasks.	
	INLIK IOI IEAT PIOCESSIIIg tasks.	

	Unit – XV	
	Part of Speech Tagging:	
Week 8	Introduction to part of speech tagging and its	
	importance in natural language understanding.	
	Unit – XVI	
	Using NLTK for part of speech tagging tasks.	
	Unit – XVII	
Week 9	Syntax Parsing:	
	Basics of syntax parsing for analyzing the	
	grammatical structure of sentences.	
	Unit – XVIII	
	Overview of parsing techniques and algorithms.	
	Unit – XIX	
	Keyword and Phrase Extraction:	
	Techniques for extracting important keywords and	
Week 10	phrases from text documents.	
	Unit – XX	
	Hands-on exercises on implementing keyword and	
	phrase extraction algorithms.	
	Unit – XXI	
	Vector Space Model and Dimensionality Reduction:	
	Introduction to the vector space model for	
XX/1-11	representing text documents as numerical vectors.	
Week 11	Unit – XXII	
	Techniques for dimensionality reduction in text data	
	using methods like Singular Value Decomposition	
	(SVD) and Principal Component Analysis (PCA).	
	Unit – XXIII	
	Topic Modeling:	
	Understanding topic modeling algorithms such as	
XX 1 10	Latent Dirichlet Allocation (LDA) for discovering	
Week 12	latent topics in text corpora.	
	Unit – XXIV	
	Implementing topic modeling using Python libraries	
	like Gensim.	
	Unit – XXV	
	Distributional Models:	
	Introduction to distributional models for	
	representing word meanings based on their	
Week 13	distributional properties in a corpus.	
	Unit – XXVI	
	Hands-on exercises on building distributional word	
	embeddings using techniques like Word2Vec and	
	GloVe.	
	Unit – XXVII	
	Text Classification:	
	Overview of text classification techniques using	
Week 14	supervised learning algorithms.	
	Unit – XXVIII	
	Using machine learning classifiers to classify text	
	Using machine rearning crassifiers to crassify text	

	documents into predefined categories.				
Week 15	Unit – XXIXSequence Labeling:Explanation of sequence labeling tasks such as named entity recognition (NER) and part of speech tagging.Unit – XXXHands-on exercises on implementing sequence labeling algorithms using machine learning and deep learning approaches.				
Week 16	Deep Learning Introduction to language proce tagging. Unit – T Building and tr	Unit – XXXI Deep Learning for NLP: POS Tagging: Introduction to deep learning techniques for natural language processing, with a focus on part of speech tagging. Unit – XXXII Building and training deep learning models for POS tagging using frameworks like TensorFlow or			
		Textbooks a	nd Reading Material		
				ge processing with Python: lia, Inc.".	
analy: Suggestee 1. Jurafs Natur	S., Klein, E., & zing text with the r d Readings ky, D., & Martin,	natural language J. H. (1999). Sp	e toolkit. "O'Reilly Med beech and Language Pr		
analy: Suggestee 1. Jurafs Natur	S., Klein, E., & zing text with the r d Readings ky, D., & Martin, al Language Pr	natural language J. H. (1999). Sp cocessing, Comp	e toolkit. "O'Reilly Med beech and Language Pr	lia, Inc.".	
analyz Suggester 1. Jurafs Natur Prenti	S., Klein, E., & zing text with the r d Readings ky, D., & Martin, al Language Pr ce Hall.	natural language J. H. (1999). Sp cocessing, Comp Teaching I includes seminars	e toolkit. "O'Reilly Med beech and Language Pr putational Linguistics Learning Strategies	lia, Inc.".	
analyz Suggester 1. Jurafs Natur Prenti	S., Klein, E., & zing text with the r d Readings ky, D., & Martin, <i>al Language Pr</i> ce Hall. ure method, which sed where necessar	natural language J. H. (1999). Sp cocessing, Comp Teaching I includes seminars y)	e toolkit. "O'Reilly Med beech and Language Pr putational Linguistics Learning Strategies	dia, Inc.". rocessing: An Introduction to , and Speech Recognition. ts and projects. (Audio-visual	
analy: Suggester 1. Jurafs <i>Natur</i> Prenti Class Lect tools are u	S., Klein, E., & zing text with the r d Readings ky, D., & Martin, <i>al Language Pr</i> ce Hall. ure method, which sed where necessar	natural language J. H. (1999). Sp cocessing, Comp Teaching I includes seminars y) gnments: Types	e toolkit. "O'Reilly Med beech and Language Pr putational Linguistics Learning Strategies , discussions, assignment	dia, Inc.". rocessing: An Introduction to , and Speech Recognition. ts and projects. (Audio-visual	
analy: Suggester 1. Jurafs <i>Natur</i> Prenti Class Lect tools are u	S., Klein, E., & zing text with the r d Readings ky, D., & Martin, al Language Pr ce Hall. ure method, which sed where necessar Assig	natural language J. H. (1999). Sp cocessing, Comp Teaching I includes seminars y) gnments: Types spective teacher.	e toolkit. "O'Reilly Med beech and Language Pr putational Linguistics Learning Strategies , discussions, assignment	dia, Inc.". rocessing: An Introduction to , and Speech Recognition. ts and projects. (Audio-visual	
analy: Suggester 1. Jurafs <i>Natur</i> Prenti Class Lect tools are u	S., Klein, E., & zing text with the r d Readings ky, D., & Martin, al Language Pr ce Hall. ure method, which sed where necessar Assig	natural language J. H. (1999). Sp cocessing, Comp Teaching I includes seminars y) gnments: Types spective teacher.	e toolkit. "O'Reilly Med peech and Language Pr putational Linguistics, Learning Strategies a, discussions, assignment and Number with Ca ssessment	dia, Inc.". rocessing: An Introduction to , and Speech Recognition. ts and projects. (Audio-visual	
analy: Suggestee 1. Jurafs Natur Prenti Class Lect tools are u According	S., Klein, E., & zing text with the r d Readings ky, D., & Martin, al Language Pr ce Hall. ure method, which sed where necessar Assig to the choice of res	natural language J. H. (1999). Sp cocessing, Comp Teaching I includes seminars y) gnments: Types spective teacher. A	e toolkit. "O'Reilly Med beech and Language Productional Linguistics, Learning Strategies , discussions, assignment and Number with Ca ssessment	dia, Inc.". rocessing: An Introduction to , and Speech Recognition. ts and projects. (Audio-visual lendar	

3.	Final Assessment	40%	It takes place at the end of the semester. It is mostly in the form of a test, but owing to the nature of the course the teacher may assess their students based on term paper, research proposal development, field work and report writing etc.
----	---------------------	-----	---