

<b>Programme</b>	BS Computational Statistics and Data Analytics	<b>Course Code</b>	CSTA-408	<b>Credit Hours</b>	3
<b>Course Title</b>	Natural Language Processing				
<b>Course Introduction</b>					
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					
<b>Learning Outcomes</b>					
By the end of this course, students will be able to:					
<ol style="list-style-type: none"> <li>1. Learn about basic NLP problems, tasks and methods.</li> <li>2. Master basic programming tools for NLP.</li> <li>3. Implement a simple NLP system.</li> <li>4. Define a NLP problem and find a suitable solution to it.</li> </ol>					
<b>Course Content</b>				<b>Assignments/Readings</b>	
<b>Week 1</b>	<b>Unit – I</b> Natural Language and Formal Language: Introduction to natural language and formal language theory.				
	<b>Unit – II</b> Understanding the distinction between natural languages and formal languages.				
<b>Week 2</b>	<b>Unit – III</b> Regular Expressions and Finite State Automata: Basics of regular expressions for pattern matching and text processing.				
	<b>Unit – IV</b> Overview of finite state automata and their applications in language processing.				
<b>Week 3</b>	<b>Unit – V</b> N-grams and Language Models: Explanation of N-grams as consecutive sequences of N items (usually words) in text.				
	<b>Unit – VI</b> Introduction to language models for predicting the likelihood of word sequences.				
<b>Week 4</b>	<b>Unit – VII</b> An Introduction to Programming in Python: Overview of Python programming language, its syntax, and basic concepts.				
	<b>Unit – VIII</b> Setting up Python development environment and writing simple Python programs.				
<b>Week 5</b>	<b>Unit – IX</b> Variables, Numbers, Strings, Arrays, Dictionaries in Python:				

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

	Latent Dirichlet Allocation (LDA) for discovering latent topics in text corpora.	
	<b>Unit – XXIV</b> Implementing topic modeling using Python libraries like Gensim.	
<b>Week 13</b>	<b>Unit – XXV</b> Distributional Models: Introduction to distributional models for representing word meanings based on their distributional properties in a corpus.	
	<b>Unit – XXVI</b> Hands-on exercises on building distributional word embeddings using techniques like Word2Vec and GloVe.	
<b>Week 14</b>	<b>Unit – XXVII</b> Text Classification: Overview of text classification techniques using supervised learning algorithms.	
	<b>Unit – XXVIII</b> Using machine learning classifiers to classify text documents into predefined categories.	
<b>Week 15</b>	<b>Unit – XXIX</b> Sequence Labeling: Explanation of sequence labeling tasks such as named entity recognition (NER) and part of speech tagging.	
	<b>Unit – XXX</b> Hands-on exercises on implementing sequence labeling algorithms using machine learning and deep learning approaches.	
<b>Week 16</b>	<b>Unit – XXXI</b> Deep Learning for NLP: POS Tagging: Introduction to deep learning techniques for natural language processing, with a focus on part of speech tagging.	
	<b>Unit – XXXII</b> Building and training deep learning models for POS tagging using frameworks like TensorFlow or PyTorch.	
<b>Textbooks and Reading Material</b>		
<b>Text Book</b>		
1. Bird, S., Klein, E., & Loper, E. (2009). <i>Natural language processing with Python: analyzing text with the natural language toolkit</i> . "O'Reilly Media, Inc."		
<b>Suggested Readings</b>		
1. Jurafsky, D., & Martin, J. H. (1999). <i>Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition</i> . Prentice Hall.		

<b>Teaching Learning Strategies</b>			
Class Lecture method, which includes seminars, discussions, assignments and projects. (Audio-visual tools are used where necessary)			
<b>Assignments: Types and Number with Calendar</b>			
According to the choice of respective teacher.			
<b>Assessment</b>			
<b>Sr. No.</b>	<b>Elements</b>	<b>Weightage</b>	<b>Details</b>
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.

<b>Programme</b>	BS Computational Statistics and Data Analytics	<b>Course Code</b>	CSTA-408	<b>Credit Hours</b>	3
<b>Course Title</b>	Natural Language Processing				
<b>Course Introduction</b>					
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					
<b>Learning Outcomes</b>					
By the end of this course, students will be able to:					
5. Learn about basic NLP problems, tasks and methods.					
6. Master basic programming tools for NLP.					
7. Implement a simple NLP system.					
8. Define a NLP problem and find a suitable solution to it.					
<b>Course Content</b>					<b>Assignments/Readings</b>
<b>Week 1</b>	<b>Unit – I</b>				

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

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

	documents into predefined categories.		
<b>Week 15</b>	<b>Unit – XXIX</b> Sequence Labeling: Explanation of sequence labeling tasks such as named entity recognition (NER) and part of speech tagging.		
	<b>Unit – XXX</b> Hands-on exercises on implementing sequence labeling algorithms using machine learning and deep learning approaches.		
<b>Week 16</b>	<b>Unit – XXXI</b> Deep Learning for NLP: POS Tagging: Introduction to deep learning techniques for natural language processing, with a focus on part of speech tagging.		
	<b>Unit – XXXII</b> Building and training deep learning models for POS tagging using frameworks like TensorFlow or PyTorch.		
<b>Textbooks and Reading Material</b>			
<b>Text Book</b>			
1. Bird, S., Klein, E., & Loper, E. (2009). <i>Natural language processing with Python: analyzing text with the natural language toolkit</i> . "O'Reilly Media, Inc."			
<b>Suggested Readings</b>			
1. Jurafsky, D., & Martin, J. H. (1999). <i>Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition</i> . Prentice Hall.			
<b>Teaching Learning Strategies</b>			
Class Lecture method, which includes seminars, discussions, assignments and projects. (Audio-visual tools are used where necessary)			
<b>Assignments: Types and Number with Calendar</b>			
According to the choice of respective teacher.			
<b>Assessment</b>			
<b>Sr. No.</b>	<b>Elements</b>	<b>Weightage</b>	<b>Details</b>
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.
----	------------------	-----	---