FACULTY OF ENGINEERING AND TECHNOLOGY

Effective from Academic Batch: 2022-23

Programme: **M.TECH. ARTIFICIAL INTELLIGENCE**

Semester: **II**

Course Code: **202340201**

Course Title: **Data Analytics with Natural Language Processing**

Course Group: **Programme Elective-IV**

Course Objectives:

This course provides a unique opportunity for you to learn key components of text mining and analytics aided by the real world datasets. NLP attempts to interact with humans and human texts via language. Problems in the domain include analysing texts to discover structures and to make decisions. Translating from one language to another. Interacting with humans in dialogue systems or cooperative tasks.

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	50/20	50/20	25/10	25/10	150/60

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

Detailed Syllabus:

Sr.	Contents	Hours
1	Data Analytics: Overview, Dealing with different types of Data, Data visualization for Decision making, Text Analytics, Linguistics, Language Syntax and Structure	4
2	Fundamentals of Natural Language Processing : Ambiguity and uncertainty in language, Models and Algorithms, Regular Expressions, Finite State Automata, Morphology, Morphological Parsing, N-grams Models.	6
3	Text Processing: Text Tokenization – Sentence, Word Text Normalization - Cleaning, Tokenizing, Removing Special Characters, Expanding Contractions, Case Conversions, Removing Stop words, Correcting Words, Stemming, Lemmatization Understanding Text Structure – Parts of Speech tagging, Shallow Parsing.	6

Opp. Shastri Maidan, Beside BVM College, Vallabh Vidyanagar, Dist: Anand, Gujarat - 388120
(O): 02692-238001 | Email: adminoffice@cvmu.edu.in | www.cvmu.edu.in

4	Classification for Text Analysis : Text Classification - Identifying Classification Problems, Classifier Models Feature Extraction – Bag of words Model, TF-IDF Model, Advanced word Vectorization Models, Classification Algorithms, Evaluating Classification Models, Applications.	8
5	Clustering for Text Similarity: Unsupervised Learning on Text Clustering by Document Similarity – Distance Metrics, Partitive Clustering, Hierarchical Clustering, Analyzing Document Similarity, Document Clustering.	8
6	Semantic and Sentiment Analysis: Semantic Analytics, Word Sense Disambiguation, Named Entity Recognition, Analyzing Semantic Representation, Sentiment Analysis	8

List of Practicals / Tutorials:

1	Introduction to python libraries for feature extraction and NLP.
2	Split the text sentence/paragraph into a list of words.
3	Tokenize words, sentence wise.
4	Remove a regex pattern from the input text.
5	Perform lemmatization and stemming.
6	Perform POS tagging annotation on input text.
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 - Identify the sentiment of tweets

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.

Supplementary learning Material:	
1	https://nptel.ac.in/courses/106/105/106105158/

Pedagogy:

- Direct classroom teaching
- Audio Visual presentations/demonstrations
- Assignments/Quiz
- Continuous assessment

Opp. Shastri Maidan, Beside BVM College, Vallabh Vidyanagar, Dist: Anand, Gujarat - 388120
(O): 02692-238001 | Email: adminoffice@cvmu.edu.in | www.cvmu.edu.in

- Interactive methods
- Seminar/Poster Presentation
- Industrial/ Field visits
- Course Projects

Internal Evaluation:
The internal evaluation comprised of written exam (40% weightage) along with combination of various components such as Certification courses, Assignments, Mini Project, Simulation, Model making, Case study, Group activity, Seminar, Poster Presentation, Unit test, Quiz, Class Participation, Attendance, Achievements etc. where individual component weightage should not exceed 20%.

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

Distribution of Theory Marks in %						**R**: Remembering; **U**: Understanding; **A**: Applying; **N**: Analyzing; **E**: Evaluating; **C**: Creating
R	**U**	**A**	**N**	**E**	**C**	
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 approaches to syntax and semantics in NLP.	20%
CO-2	Understand approaches to discourse, generation, dialogue and summarization within NLP.	30%
CO-3	Understand current methods for statistical approaches to machine translation.	30%
CO-4	To introduce basic mathematical models and methods used in NLP Applications to formulate computational solutions.	20%

Curriculum Revision:	
Version:	2.0
Drafted on (Month-Year):	June-2022
Last Reviewed on (Month-Year):	-
Next Review on (Month-Year):	June-2025

Opp. Shastri Maidan, Beside BVM College, Vallabh Vidyanagar, Dist: Anand, Gujarat - 388120
(O): 02692-238001 | Email: adminoffice@cvmu.edu.in | www.cvmu.edu.in