THE INTERNATIONAL QUALIFICATIONS

AND ASSESSMENT CENTRE (IQAC)

«Approved»

2025 y. «16» ____

<table>
<tr><td>Programme</td><td colspan="3">Level 7 Diploma in Data Science</td></tr>
<tr><td>Unit Number/ Unit Title</td><td colspan="3">UNIT 1 NATURAL LANGUAGE PROCESSING AND TEXT ANALYTICS</td></tr>
<tr><td>Cohort Code:</td><td colspan="3">L07NLTA-U1</td></tr>
<tr><td>Unit Level</td><td colspan="3">Level 7</td></tr>
<tr><td>Total GLH</td><td colspan="3">Total qualification time 200/ Total Guided learning hours 90/ Self-guided learning hours 110</td></tr>
<tr><td>Credits</td><td colspan="3">20 CATS/ 10 ECTS</td></tr>
<tr><td>Lecturer</td><td colspan="3"></td></tr>
<tr><td>Start Date</td><td></td><td>End Date</td><td></td></tr>
</table>

Unit Aims	This unit aims to develop students' expertise in Natural Language Processing (NLP) and Text Analytics. It explores key algorithms, models, and applications for processing and analyzing unstructured text data. Students will learn to build and evaluate NLP models, apply linguistic preprocessing techniques, and extract insights from real-world textual datasets.
Differentiation Strategies *(e.g. planned activities or support for individual learners according to their needs)*	The total number of students to be in the lesson is approximately 20. This is a multicultural group of students predominantly between the ages of 24 – 45, with numerous ethnic, gender, and creed background. These are UK academic level 5 students; hence it is assumed that they have practical, theoretical, or technological knowledge and understanding of a subject or field of work to find ways forward in broadly defined, complex contexts. These students must be able to generate information, evaluate, synthesise the use information from a variety of sources. Various approaches to addressing the various identified students needs will be adopted throughout the lesson. Such will include:- 1. Progressive tasks 2. Digital resources

	3. Verbal support 4. Variable outcomes 5. Collaborative learning 6. Ongoing assessment 7. Flexible-pace learning
Equality & Diversity	Variety of teaching techniques will be employed to ensure that the needs of each individual learner are met.
Safeguarding & Prevent	Safeguarding policies and the Prevent duty are strictly observed to ensure the safety, well-being, and inclusivity of all students and staff.
Health & Safety	SIRM H&S policies will be maintained.
Learning Resources	**Teaching and Learning Materials**
	• Jurafsky, D., & Martin, J. H. (2020). Speech and Language Processing. Pearson. • Bird, S., Klein, E., & Loper, E. (2009). Natural Language Processing with Python. O'Reilly. • Cambria, E., & White, B. (2014). Jumping NLP Curves. IEEE Computational Intelligence Magazine. • Eisenstein, J. (2019). Introduction to Natural Language Processing. MIT Press.

Learning Outcome	Assessment Criteria
LO1. 1. Apply core NLP techniques for text processing.	**Written Report:** 1.1 Implement text pre-processing methods (tokenization, stemming, lemmatization). 1.2 Demonstrate understanding of POS tagging and syntactic parsing.
LO2. 2. Analyze and interpret textual data using machine learning.	**Practical Project:** 2.1 Train and evaluate text classification models. 2.2 Use vectorisation techniques (TF-IDF, word embeddings).
LO3. 3. Explore applications of sentiment analysis and topic modelling.	**Presentation**: 3.1 Apply sentiment analysis on social media/textual data. 3.2 Use LDA or NMF for topic modelling
LO4. 4. Evaluate challenges in multilingual and ethical NLP systems.	**Case Study**: 4.1 Assess bias and fairness in NLP models. 4.2 Propose solutions for multilingual text processing.

No	Learning Outcome / Topic	Learning and Teaching Activities	Which assessment criteria does the session relate to?	Day/month/year/ signature
1.	**Introduction to NLP**	**Introduction to NLP** Applications, challenges, and NLP pipeline overview	LO1: Core NLP Techniques	
2.	**Text Preprocessing**	**Text Preprocessing** Tokenization (NLTK, spaCy), stopword removal	LO1: Core NLP Techniques	
3.	**Stemming & Lemmatization**	**Stemming & Lemmatization** Porter vs. Snowball stemmer, WordNet lemmatization	LO1: Core NLP Techniques	
4.	**Part-of-Speech (POS) Tagging**	**Part-of-Speech (POS) Tagging** Rule-based vs. statistical tagging (spaCy, HunPos)	LO1: Core NLP Techniques	
5.	**Syntactic Parsing**	**Syntactic Parsing** Dependency parsing, constituency parsing (Stanford Parser)	LO1: Core NLP Techniques	
6.	**Bag-of-Words (BoW) Model**	**Bag-of-Words (BoW) Model** CountVectorizer, n-grams	LO2: Text Analysis with ML	
7.	**TF-IDF Vectorization**	**TF-IDF Vectorization** Theory, sklearn implementation, limitations	LO2: Text Analysis with ML	
8.	Half-Term Exam	- Review of LO1 topics - Practice questions and mock assessment - **Half-term assessment** based on LO1 (theory)	LO1 LO2	
9.	**Word Embeddings**	**Word Embeddings** Word2Vec (CBOW, Skip-gram), GloVe	LO2: Text Analysis with ML	
10.	**Text Classification**	**Text Classification** Naïve Bayes, SVM, logistic regression for sentiment analysis	LO2: Text Analysis with ML	
11.	**Deep Learning for NLP**	**Deep Learning for NLP** CNNs for text, RNNs/LSTMs for sequence modeling	LO2: Text Analysis with ML	

12.	**Sentiment Analysis**	**Sentiment Analysis** VADER, fine-tuning BERT for sentiment	LO3: Advanced NLP Applications	
13.	**Topic Modeling**	**Topic Modeling** LDA (Latent Dirichlet Allocation), NMF (Non-negative Matrix Factorization)	LO3: Advanced NLP Applications	
14.	Final Exam Preparation & Review	- Comprehensive review of all learning outcomes - Practice questions and revision of key topics		
15.	Final Exam	- **Final-term assessment** covering all learning outcomes (theory and practical elements)		
16.	Feedback & Reflection	- Review of final exam - Individual feedback on performance - Reflective discussion on key learning points		
17.	**Named Entity Recognition (NER)**	**Named Entity Recognition (NER)** spaCy, BERT-based NER	LO3: Advanced NLP Applications	
18.	**Text Summarization**	**Text Summarization** Extractive (TextRank) vs. abstractive (T5, BART)	LO3: Advanced NLP Applications	
19.	**Question Answering Systems**	**Question Answering Systems** SQuAD dataset, BERT for QA	LO3: Advanced NLP Applications	
20.	**Multilingual NLP Challenges**	**Multilingual NLP Challenges** Low-resource languages, transfer learning (mBERT, XLM-R)	LO4: Multilingual & Ethical NLP	
21.	**Bias in NLP Models**	**Bias in NLP Models** Gender/racial bias in word embeddings (Bolukbasi et al.)	LO4: Multilingual & Ethical NLP	
22.	**Fairness & Mitigation Strategies**	**Fairness & Mitigation Strategies** Debiasing techniques, adversarial training	LO4: Multilingual & Ethical NLP	
23.	Half-Term Exam	**Project** Solve a real-world NLP problem (e.g., fake news detection)		

24.	**Ethical Considerations**	**Ethical Considerations** Misinformation, deepfake text, consent in data collection	LO4: Multilingual & Ethical NLP	
25.	**Case Study: Bias in Hiring Tools**	**Case Study: Bias in Hiring Tools** Amazon's biased recruitment AI	LO4: Multilingual & Ethical NLP	
26.	**Building an NLP Pipeline**	**Building an NLP Pipeline** End-to-end project: Data → Preprocessing → Model → Deployment	LO5: Practical Implementation	
27.	**Transformer Models**	**Transformer Models** BERT, GPT, T5 (Hugging Face `transformers`)	LO5: Practical Implementation	
28.	**Model Interpretability Deploying NLP Models**	**Model Interpretability** SHAP, LIME for NLP models **Deploying NLP Models** Flask API, Docker, Hugging Face Spaces	LO5: Practical Implementation	
29.	Final Exam Preparation & Review	LO1, LO2, LO3, LO4	LO1, LO2, LO3, LO4	
30.	Final Exam		LO1, LO2, LO3, LO4	