Introducing Module 7 – Natural Language Processing

Pragya Paudyal &
Filippo Cavallari

Data Science Lecturers

13 May 2024

Data Science Campus

Learning Outcomes

- Describe what is meant by Natural Language Processing and understand what the benefits and challenges of such approaches are.
- Understand the components of Regular Expressions and implement them to match string patterns in unstructured text data.
- Apply pre-processing methods such as tokenisation to prepare unstructured text data for language analysis and modelling.
- Understand the use of corpora for text representation and analyse their structure with algorithms such as Bag of Words.
- Create structures such as the Document Term Matrix (DTM) and perform simple language modelling with Sentiment Analysis (in R), Word Embeddings and N-grams (in Python).

Day 1

Time	Session
09:30 – 09:45	Welcome and Learning Objectives
09:45 – 11:15	Introduction to Natural Language Processing
11:15 – 12:00	Independent Learning: Python and R Chapter 1
12:00 – 13:00	**Lunch**
13:00 – 15:30	Regex Workshop
15:30 – 16:00	Independent Learning: Python and R Chapter 2

Day 2

Time	Session
09:30 – 09:45	Morning Check-in
09:45 – 11:30	R/Python Workshop
11:30 – 12:00	Independent Learning: Python and R Chapter 3
12:00 – 13:00	**Lunch**
13:00 – 14:50	Independent Learning: Python and R Chapters 3 & 4
15:00 – 16:00	Case Study

Day 3

Time	Session
09:30 – 09:45	Morning Check-in
09:45 – 10:45	Introduction to Large Language Models (LLMs)
10:45 – 12:00	Independent Learning: Python and R Chapter 5
12:00 - 13:00	**Lunch**
13:00 – 14:00	Case Study: LLMs • Parli AI • Ollama LLMs
14:00 – 14:45	Independent Learning: Python and R Chapter 5
14:45 - 15:30	**Fun:** Kahoot Quiz
15:30 – 16:00	Wrap-up & survey
16:00 – 16:30	Graduate Representatives: Town Hall