Course Title: Advanced Natural Language Processing in Scientific Publications Analysis

Course Duration: 16 hours (with lab)

Course Description:

This course has been designed to equip participants with a profound comprehension of and practical expertise in **analysing scientific publications** through the application of **advanced data analysis** methods. It comprises five distinct sections, starting with the formulation of a **search strategy** for accessing scientific publications. The course encompasses the utilization of **metadata** from scientific publications for conducting literature review analyses, as well as delving into more advanced **Natural Language Processing** (NLP) techniques for scoping review analyses. Furthermore, it presents an overview of both supervised and unsupervised NLP methodologies for scrutinizing scientific articles. To conclude, the course incorporates **hands-on applications** and **case studies**, enabling students to apply these techniques within their respective fields of expertise.

Main Topics

Topic 1: Search Strategy on Scientific Publications database

1. Overview of the course
2. Importance of search strategy for scientific research
3. Scientific Publications Database (Scopus, Google Scholar, Web of Science)
4. Definitions and concepts in query design (OR, AND operators)
5. Evaluation Metrics (Precision and Recall)
6. Hands-on exercises and case studies

Topic 2: Metadata and Network analysis from Scientific Papers

1. Metadata analysis in Scientific Papers (co-authorship analysis, keywords analysis, trend analysis)
2. Applications of network analysis in scientific research
3. Network Theory: Definitions, Measures and Visualization
4. Network analysis as a tool for extracting insights in R

Topic 3: Natural Language Processing for Scientific Publications Analysis: Supervised Methods

1. Introduction to Scoping Review: Definition and Process
2. Conducting a scoping review with Natural Language Processing in R
3. Reporting and presenting scoping review results in R

Topic 4: Natural Language Processing for Scientific Publications Analysis: Un-supervised Methods

1. Introduction to Unsupervised Analysis of Scientific Papers
2. Clustering and topic modeling of scientific papers with BERTopic
3. Reporting and presenting BERTopic results

Topic 5: Scientific Publications Analysis – Hands-on lab

1. Metadata analysis with igraph in RStudio, Gephi for network visualization
2. Case study on supervised analysis
3. Case study on unsupervised analysis

References

Chiarello, F., Belingheri, P., & Fantoni, G. (2021). Data science for engineering design: State of the art and future directions. *Computers in Industry*, 129, 103447.

Belingheri, P., Chiarello, F., Fronzetti Colladon, A., & Rovelli, P. (2021). Twenty years of gender equality research: A scoping review based on a new semantic indicator. *Plos one*, 16(9), e0256474.

Chiarello, F., Gastaldi, L., & Martini, A. (2023). Design and implementation of a text mining-based tool to support scoping reviews. *International Journal of Technology Management*, 91(3-4), 147-161.