Program(s): **(Executive) MBA DD & Data Science certificate program**

Syllabus for Module: **Natural Language Processing & AI**

Date:	**Lecturer:**	**ECTS:**	**Language:**
December 11-12, 2024	Daniel Dan	2	English

Course Description

This course offers a comprehensive introduction to Natural Language Processing (NLP) and its applications in modern technology. Students will begin by exploring classical NLP techniques, including topic modeling, classification, sentiment analysis, and named entity recognition (NER), code will be provided in R. Through these foundational methods, students will gain insights into how machines process and analyze human language.

The course then transitions into more advanced NLP concepts, such as transformation of words into embeddings transformers, and the architecture of chatbots. Students will learn the fundamentals of chatbot implementation, covering each component, including response generation strategies and tools like Retrieval-Augmented Generation Systems (RAGS). Through guided instruction, they will explore the practicalities of fine-tuning models to customize a chatbot and employing LLMs to accomplis this. In the final segment, students will actively build and personalize their own chatbots. Using Python, each participant will use provided code to integrate their unique datasets, ultimately creating and fine-tuning a chatbot tailored to specific needs and contexts.

Objectives and Learning Outcomes

By the end of the module students will have learned:

- The technicalities and operations of transforming text into numbers
- Theoretical knowledge and practical skills in NLP
- Hands-on experience in building, fine-tuning, and creating personalized chatbots

Knowledge and understanding

After completing this module the students will have:

- Have the R/Python knowledge to perform NLP tasks
- A strong understanding of implementing NLP in industry
- Know how to build and deploy a personalized chatbot

Cognitive and subject specific skills

After completing this module the students will have the skills:

- Reuse and recycle R/Python code for generic NLP tasks
- Know the dos and donts in a NLP task
- Easily setup a chatbot for their daily routines or their company

Key skills

After completing this module the students will:

- Know how to use programming languages for NLP tasks
- Immediately delpoy chatbots
- Transform an NLP idea into a product

Course Structure and Teaching Methods

Day 1, Before noon and after noon, lecture. In the afternoon, group formation, choice of datasets to employ.

Day 2, Group work, choice of a task, group presentations.

Transversal Themes and Current Discourse

Contemporary issues highlighted in this module:

- Ethical issues, algorithmic bias and fairness (mentioned)
- The chatbot and LLMs as new team members
- Limits of datasets or LLMs
- Generative Artificial Intelligence

Reading Material

Required readings

- Silge, J. (2017). Text mining with R: A tidy approach. O'Reilly Media, Inc.Available at: https://www.tidytextmining.com/
- Jurafsky, D., & Martin, J. H. Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. Available at: https://web.stanford.edu/~jurafsky/slp3/

Additional readings

- Ciesla, R. (2024). The Book of Chatbots: From ELIZA to ChatGPT. Springer Nature.
- Kwartler, T. (2017). Text mining in practice with R. John Wiley & Sons.

Recommended links and websites

- A series of free, online R books. Available at: https://github.com/jl5000/r_ebooks
- A series of free, online Python books. Available at: https://github.com/pamoroso/free-python-books

Guidelines Regarding AI Tools

AI tools, such as ChatGPT and others, are allowed for any assignment with proper citatation.

Pre-Module

Reading assignments/material

- See required readings above

Pre-module Assignment (Individual Activity)

Assigment #1

Explore a Natural Language Processing (NLP) task you would be interested in developing. Write a short essay (around 2 pages, 12pt font, Times New Roman) where you describe this task, why it is important for you and how you might go about achieving it. This assignment is meant to be exploratory, so do not worry if you are new to NLP concept, focus more on brainstorming and explaining your ideas.

Hints: Think of a task where a computer might analyze or use text data to do something meaningful, for example, analyze customer feedback, find important keywords in news articles or social media posts, group documents by topics (e.g., sports, politics, entertainment), summarizing long articles into shorter versions, translate text from one language to another, or come up with you own idea based on your own interests!

Some questions you might ask yourself: What do I want to accomplish and why? Who would benefit from this task? Why is this of interest? What data to I have? What are the basic steps? What programming language should I use?

Describe what a successful outcome would look like? How can these results be used and what actionable insights can I draw?

Organize your thoughts with clear sections for each part of the assignment.

Deadline for pre-module assignment and process of delivery:

December 11, 2024, 09:00 a.m. (CET) via Moodle

All student work is checked for both plagiarism and the use of AI upon submission.

Please submit your assignment as a PDF file onto Moodle, indicate your names in the document and name your document "SURNAME_PRE".

Core-Module

Reading assignments/material

The required reading is for you to get familiar to the material taught in class. You do not need to learn by heart anything. The additional readings are useful for your learning journey.

Detailed Schedule

Day 1

08:30 – 10:30 NLP introduction to useful concepts, such as Topic Modeling, Sentiment Analysis, Named Entity Recognition

10:30 – 10:45 *Coffee break*

10:45 – 12:15 Word Embeddings and Transformers
12:15 – 13:15 *Lunch time*
13:15 – 15:15 Introduction to chatbots
15:15 – 15:30 *Coffee break*
15:30 – 17:30 Chatbot deployment, training, and fine-tuning.

Day 2

08:30 – 10:30	Code your own chatbot
10:30 – 10:45	*Coffee break*
10:45 – 12:15	Chatbot presentation, three slides, done in groups.
12:15 – 13:15	*Lunch time*

Post-Module

Reading assignments/material

Post-module assignment (Group Activity)

Prepare a series of approximately 15 slides to present and describe your chatbot. Avoid including code on the slides unless it is essential to explain a specific feature or function. Design the slides as a pitch to showcase and sell your chatbot.

Please **specify** in your last slide the contribution of each team member to the final document/project.

Deadline for post-module assignment and process of delivery:

December 31, 2024, 09.00 a.m. (CET), via Moodle

All student work is checked for both plagiarism and the use of AI upon submission.

Please submit your assignment as a PDF file onto Moodle, indicate your names in the document and name your document "GROUP NAME_POST".

Assessment

Pre-module:	30 %
Core-module:	30 %
Post-module:	40 %
	=100%

Instructor and Contact Information

Daniel Dan

Organization: Institute for Marketing and Customer Analytics | Vienna University of Economics and Business

Position: External Lecturer

E-mail: daniel.dan@wu.ac.at

Short CV of Instructor and Contact Information

Daniel Dan is employed as an External Lecturer at the Institute for Marketing and Customer Analytics. He has a University degree in Computer Science and got his PhD in Managerial and Actuarial Science from the University of Udine, in Italy. He is particularly involved in text mining methods and sentiment analysis of large text corpora. He developed interest in Computer Science at the University of Udine, having spent the junior and senior years studying and working, applying the knowledge gained to various computer programming projects. After the graduation he was linked to the Academic world through various contracts, cooperations and teaching positions. He spent the first year of his PhD programme studying Statistics at the University of Padua, Italy. In the subsequent years he dedicated his energies to practicing and applying various machine learning techniques. Currently he is focused on Text Mining algorithms applied to several disciplines. He was and is involved in several European projects that deal with Data Spaces, Information Overload, applied AI technologies, and Carrying Capacities for Tourism.