

Tutorial overview

The tutorial will offer both a theoretical understanding and practical skills of working with the method called "topic modelling". Using this method developed on the intersection of machine learning and natural language processing participants will learn new ways on how to: 1) cluster textual data into meaningful topics, 2) relate this information to other characteristics of the texts and 3) discuss potential ways how to develop a storyline around this type of analysis.

This method can be applied for example on responses to open-ended survey questions to elicit opinions, to research articles for large-scale literature reviews, to posts in social networks like Twitter and many more. Using R software a few examples of applying the method will be discussed in detail with the participants.

The main objectives of this tutorial are:

1. Provide an overview of textual data analysis techniques.
2. Explain in detail the method called structural topic modelling (STM).
3. Gain practical skills in applying STM to different data.

At the end of the tutorial its participants will be able to:

- Distinguish between different alternative methods of textual data analysis.
- Explain intuition of topic modelling.
- Show critical thinking on the strength and limits of STM.
- Perform a project using STM.

Before the tutorial.

- Why topic modelling? Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, 55(4), 77-84. <https://dl.acm.org/doi/pdf/10.1145/2133806.2133826>
- Roberts, M.E., Stewart, B.M., Tingley, D. 2019. stm: An R Package for Structural Topic Models. *Journal of Statistical Software* 91 (2), <https://doi.org/10.18637/jss.v091.i02>
- Introduction to R and RStudio (in case you are not familiar)
 - RStudio tutorial from DataCamp <https://www.datacamp.com/tutorial/r-studio-tutorial>
 - R programming tutorial <https://www.youtube.com/watch?v=V8eKsto3Ug>
 - More resources at <https://education.rstudio.com/learn/beginner/>, in particular <https://posit.co/resources/videos/a-gentle-introduction-to-tidy-statistics-in-r/>

Organizational:

- All tutorial materials at this [link](#)
- Lecture, slides and code comments in English
- Participants are recommended to
 - Bring computer with R & Rstudio installed (if you want to practice)
 - Read the materials mentioned in this document

Further reading list (applications)

1. Savin, I., Chukavina, K. & Pushkarev, A. Topic-based classification and identification of global trends for startup companies. *Small Bus Econ* 60, 659–689 (2023). <https://doi.org/10.1007/s11187-022-00609-6>
2. Savin, I., Coeurderoy R. (2024) Reviewing and classifying scientific contributions from ESCP Business School https://www.researchgate.net/publication/381377136_Reviewing_and_classifying_scientific_contributions_from_ESCP_Business_School
3. Savin, I., Drews, S., & van den Bergh, J. (2024). Carbon pricing—perceived strengths, weaknesses and knowledge gaps according to a global expert survey. *Environmental Research Letters*, 19(2), 024014. <https://doi.org/10.1088/1748-9326/ad1c1c>