How to Make Use of Machine Learning & Microsimulation in Official Statistics
3 ECTS credits
Winter 2022/23

Instructors

Hanna Brenzel, Federal Statistical Office Germany, hb@hanna-brenzel.de
Piet Daas, Statistics Netherlands, pjh.daas@cbs.nl
Hariolf Merkle, Federal Statistical Office Germany, hariolf.merkle@gmail.com
Marco Puts, Statistics Netherlands, m.puts@cbs.nl

Video lecture by

Hanna Brenzel, Federal Statistical Office Germany
Piet Daas, Statistics Netherlands
Hariolf Merkle, Federal Statistical Office Germany
Marco Puts, Statistics Netherlands

Short Course Description

The course gives an overview about advanced topics in official statistics such as Big Data, machine learning, and microsimulations. The benefits and downsides of using Big Data as a data source for official statistics production are discussed and examples of its use are given, including machine learning applications. In addition, the course provides insights into microsimulation and gives an overview of the past, the present, and the future state-of-the-art of microsimulation methods and applications within official statistics.

Course and Learning Objectives

By the end of the course, students will…

- have a basic knowledge on how to use Big Data in official statistics
- be able to list the advantages and downsides of particular Big Data sources
- be able to apply machine learning and interpret its findings from a statistics perspective
- have basic knowledge on the developments of microsimulations within and outside Germany
- be able to describe what microsimulation is and what different types of microsimulations exist
- be able to apply basic microsimulation applications

Prerequisites

Basic R knowledge is required.

Having some familiarity with the official statistics system as taught in Walter Radermacher's "The Public Good Statistics" course can be helpful.

Class Structure and Course Concept

This is an online course using a flipped classroom design. It covers the same material and content as an on-site course but runs differently. In this course, you are responsible for watching video-recorded lectures and reading the required literature for each unit prior to participating in mandatory weekly one-hour online meetings where students have the chance to discuss the materials from a unit with the instructor. Just like in an on-site course, homework will be assigned and graded.

Although this is an online course where students have more freedom in when they engage with the course materials, students are expected to spend the same amount of time overall on all activities in the course – including preparatory activities (readings, studying), in-class-activities (watching prerecorded videos, attending the live online meetings), and follow-up activities (working on assignments and exams) – as in an on-site course.

Mandatory Weekly Online Meetings

Mondays, 02:00 PM – 02:50 PM ET CET, starting November 14

Meetings will be held online through Zoom. Follow the link to the meeting sessions on the course website on https://www.elms.umd.edu/. If video participation via Internet is not possible, arrangements can be made for students to dial in and join the meetings via telephone.

In preparation for the weekly online meetings, students are expected to watch the lecture videos and read the assigned literature before the start of the meeting. In addition, students are required to post questions or comments about the materials covered in the videos or readings of the week in the forum before the meetings (deadlines for posting are 24 hours before the online meetings).

Students have the opportunity to use the Zoom meeting room set up for this course to connect with peers outside the scheduled weekly online meetings (e.g., for study groups). Students are encouraged to post the times that they will be using the room to the course website forum to avoid scheduling conflicts. Students are not required to use Zoom and can use other online meeting platforms such as Google Hangout or Skype.

Grading
Grading will be based on:

- Assignments: Examination Sheets

Long Course Description
The course gives an overview about various current topics in official statistics such as Big Data, and microsimulations. The topics are therefore independent of one another but must be considered together when it comes to current advanced topics in official statistics. The course provides an overview on the statistical use of Big Data and provides essential background knowledge to enable this. The course consists of 4 sections: one each on the use of Big Data and the use of machine learning in official statistics, and two lectures on applying microsimulation techniques. A considerable number of examples will be discussed. The Big Data section will provide a general look at the benefits and downsides of Big Data in official statistics. It also is the starting point for Big Data methodology development. In addition, the relation between Big Data analysis and the various Big Data IT environments is discussed. Next, the application of machine learning, as a subset of Artificial Intelligence methods, in the context of official statistics and Big data is discussed with particular emphasis on the statistical perspective.

The second part of the course lies on microsimulation. The planning and further development of political decisions increasingly requires the use of special simulation and calculations that go beyond the published statistical standard tables of official statistics, in order to be able to evaluate and estimate consequence of political measures. The first lecture on microsimulation provides a basic overview about the idea of microsimulation, its origin and development over time and highlights the strong relationship between the developments of microsimulation and access to individual data. Within the second lecture in microsimulation, the different types of microsimulation will be presented as well as insights to microsimulation techniques and a hands-on on their applications.

Readings
There is no text book for this course.

Required and recommended readings are provided below for each specific unit.

Course Evaluation
In an effort to improve the learning experience for students in our online courses, students will be invited to participate in an online course evaluation at the end of

the course (in addition to the standard university evaluation survey). Participation is entirely voluntary and highly appreciated.

Class Schedule
Please note that assignments and dates are subject to change. Information (e.g., articles and assignments) posted to the course website supersedes the information noted here.

Unit 1: Introduction to Big Data in official statistics
Video lecture: available Wednesday, November 02, 2022

Online meeting: Monday, November 14, 2022, 02:15 PM – 03:15 PM CET

Readings:
Daas, P.J.H., Puts, M.J.H. (2014) Big Data as a Source of Statistical Information. *The Survey Statistician* 69, 22-31. Available at: http://isi-iass.org/home/wp-content/uploads/N69-2014-01-ISSN.pdf

Ginsberg et al. (2009) Detecting influenza epidemics using search engine query data. *Nature* 457, 1012-1014. Available at: https://www.nature.com/articles/nature07634.pdf

Lazer et al. (2014) The Parable of Google Flu: Traps in Big Data Analysis. Science 343(6176), 1203-1205. Available at: https://gking.harvard.edu/files/gking/files/0314policyforumff.pdf

Recommended (optional):
Daas et al. (2015) Big Data and Official Statistics. Journal of Official Statistics 31(2), 249-262. Available at: https://content.sciendo.com/downloadpdf/journals/jos/31/2/article-p249.xml

Kitchen, R. (2014) Big Data, new epistemologies and paradigm shifts. Big Data & Society, April-June, 1-12. Available at: https://journals.sagepub.com/doi/pdf/10.1177/2053951714528481

Unit 2: Using Big Data and machine learning for official statistics
Video lecture: available Monday, November 14, 2022

Online meeting: Monday, November 21, 2022, 02:15 PM – 03:15 PM CET

Readings:

Puts, M.J.H., Daas, P.J.H. (2021) Machine Learning from the Perspective of Official Statistics. The Survey Statistician 84. pp. 12-17. Available at: http://isi-iass.org/home/wp-content/uploads/Survey_Statistician_2021_July_N84_02.pdf

Recommended (optional):
Puts et al. (2015) Finding Errors in Big Data. *Significance* 12 (3), 26-29. Available at: https://rss.onlinelibrary.wiley.com/doi/epdf/10.1111/j.1740-9713.2015.00826.x

Unit 3: Introduction to microsimulation in official statistics

Video lecture: available Monday, November 21, 2022

Online meeting: Monday, November 28, 2022, 02:15 PM – 03:15 PM CET

Readings:
Orcutt (2007) A new type of socio-economic system. International Journal of Microsimulation (Reprinted), 1(1), pp. 3-9. Available at: https://www.microsimulation.org/IJM/V1_1/IJM_1_1_2.pdf

Recommended (optional):
Li and O'Donoghue (2013) A survey of dynamic microsimulation models: uses, model structure and methodology. International Journal of Microsimulation 6(2), pp. 3-55. Available at: https://www.microsimulation.org/IJM/V6_2/2_IJM_6_2_2013_Li_Odonoghue.pdf

Rahman and Harding (2017) Small area estimation and microsimulation modelling. Chapman and Hall/CRC, London, UK, Chapter 4.

Merz (1991) Microsimulation - a survey of principles, developments and applications. International Journal of Forecasting, 7(1), pp. 77-104. Available at: https://mpra.ub.uni-muenchen.de/7232/

Unit 4: Hands-on microsimulation techniques

Video lecture: available Monday, November 28, 2022

Online meeting: Monday, December 5, 2022, 02:15 PM – 03:15 PM CET

Readings:

Dekkers, G. (2015). The simulation properties of microsimulation models with static and dynamic ageing–a brief guide into choosing one type of model over the other. *International Journal of Microsimulation*, *8*(1), 97-109. https://microsimulation.pub/articles/00113

O'Donoghue, C., & Loughrey, J. (2014). Nowcasting in microsimulation models: a methodological survey. *Journal of Artificial Societies and Social Simulation*, *17*(4), 12.
https://www.jasss.org/17/4/12/12.pdf

Recommended (optional):

Lovelace, R., Dumont, M., Ellison, R., & Založnik, M. (2017). *Spatial microsimulation with R*. Chapman and Hall/CRC.
https://spatial-microsim-book.robinlovelace.net/index.html

Tanton, R. (2014). A review of spatial microsimulation methods. *International Journal of Microsimulation*, *7*(1), 4-25.
https://microsimulation.pub/articles/00092