Advanced Methods | Big Data Analysis

Zeppelin University Friedrichshafen, April 26 & 27, 2019

Prof. Dr. Karsten Donnay
karsten.donnay@uni-konstanz.de

Course Description

This block course provides a basic introduction to big data and corresponding quantitative research methods. The objective of the course is to familiarize students with big data analysis as a tool for addressing substantive problems. The course begins with a basic introduction to big data and discusses what the analysis of these data entails, as well as associated technical, conceptual and ethical challenges. Strength and limitations of big data research are discussed in depth using real-world examples. Students then engage in case study exercises in which small groups of students develop and present a big data concept for a specific real-world case. These exercises are designed to familiarize students with the format of big data and to gain a first, hands-on experience with potential applications for large, complex data in policy-relevant settings. There are no prerequisite requirements for this course.

Overview of the Course

The first day focuses on providing a theoretical and practical introduction to big data, its analysis and associated challenges. During the second day, students then apply this knowledge in an in-depth case study and prepare a case study report and oral presentation. The course provides a conceptual overview of technical aspects of big data analysis but students are not required to complete practical programming exercises and no prior knowledge of scripting etc. is assumed.

Day 1: Fundamentals of Big Data Analysis

1. Introduction – What is Big Data?

2. Handling and Processing Big Data

3. Methodological Challenges and Problems

4. Example Applications

Day 2: Big Data Analysis in Practice

5. Case Study Session 1

6. Case Study Session 2

7. Preparation of Case Study Report and Presentation

8. Case Study Presentation

Course Schedule with Recommended Readings and Online Resources

Day 1: Fundamentals of Big Data Analysis

Session 1: Introduction – What is Big Data?
Friday, Apr. 26, 13:30-15:00

Dutcher, Jenna. (2014). What is Big Data? *UC Berkeley Data Science Blog*.
Press, Gil. (2014). 12 Big Data Definitions: What's Yours? *Forbes Blog*.
Manovich, Lev. (2012). Trending: The Promises and the Challenges of Big Social Data. *Debates in the Digital Humanities*, edited by Matthew K. Gold. The University of Minnesota Press.
Lazer, David, Alex Pentland, Lada Adamic, Sinan Aral, Albert-László Barabási, Devon Brewer, Nicholas Christakis, Noshir Contractor, James Fowler, Myron Gutmann, Tony Jebara, Gary King, Michael Macy, Deb Roy, and Marshall Van Alstyne. (2009). Computational Social Science. *Science* 323(5915): 721-723.

Session 2: Handling and Processing Big Data
Friday, Apr. 26, 15:00-16:30

Atz, Ulrich. (2013). 11 Tips on How to Handle Big Data in R. *Open Data Institute Blog*.
Lockwood, Glenn. (2014). Conceptual Overview of Map-Reduce and Hadoop. *Blog Post*.
Jacobs, Bill. (2015). Using Hadoop with R: It Depends. *Blog Post*.
Penchikala, Srini. (2015). Big Data Processing in Apache Spark – Part 1: Introduction. *InfoQ Article*.
Venkataraman, Shivaram. (2015). Announcing SparkR: R on Spark. *Databricks Blog Post*. (*R* package)

Session 3: Methodological Challenges and Problems
Friday, Apr. 26, 16:30-18:00

Bollier, David (2010). The Promise and Peril of Big Data. *The Aspen Institute Communications and Society Program*.
Cate, Fred H. (2014). The Big Data Debate. *Science* 346(6211): 818-818.
Lazer, David, Ryan Kennedy, Gary King, and Alessandro Vespignani. (2014). The Parable of Google Flu: Traps in Big Data Analysis. *Science* 343(6176): 1203-1205.
Lazer, David. (2015). The Rise of the Social Algorithm. *Science* 348(6239): 1090-1091.

Session 4: Example Applications
Friday, Apr. 26, 18:00-19:00

Viktoria Spaiser, Thomas Chadefaux, Karsten Donnay, Fabian Russmann, and Dirk Helbing. (2017). Communication Power Struggles on Social Media: A Case Study of the 2011-12 Russian Protests. *Journal of Information Technology & Politics* 14(2): 132-153.
Karsten Donnay. (2017). Big Data for Monitoring Political Instability. *International Development Policy* 8.1 (Online).
Pablo Barberá and Thomas Zeitzoff. (2018). The New Public Address System: Why Do World Leaders Adopt Social Media? *International Studies Quarterly* 62(1): 121-130.

<u>Day 2: Big Data Analysis in Practice</u>

Session 5: Case Study Session 1
Saturday, Apr. 27, 10:00-11:30

Selection of case study topics and formation of small working groups. Students engage with the cases, read through background material provided in the session and work through an initial set of questions to deepen the understanding of the case. Sample applications and data is provided to help students familiarize themselves with the cases and available (big) data.

Session 6: Case Study Session 2
Saturday, Apr. 27, 11:30-13:00

Groups are given a specific task relevant to the case in question and are expected to develop a corresponding big data concept using the knowledge gained in the course and the parameters set by the case study scenario. A set of questions that help guide through the scenarios will be provided.

Session 7: Preparation of Case Study Report and Presentation
Saturday, Apr. 27, 13:30-15:00

Each group prepares a short 5 - 10 page written report on their results and a 10 min oral presentation of their big data concept. There are no further requirements on the exact format of the report or the how the results are presented to the course (slides, flipchart etc.).

Session 8: Case Study Presentations
Saturday, Apr. 27, 15:00-16:30

Presentation of big data concept to the group and discussion of results.