KAN-CCMVV2424U Causal Data Science for Business Decision Making 2021/2022

Learning objectives

At the end of the course, students should be able to:

- Understand the crucial role of causal knowledge for data-augmented decision-making in strategic management
- Have a precise understanding of what it means to say "*correlation doesn't imply causation*"
- Critically reflect on the shortcomings of current correlation-based approaches to machine learning and AI for business analytics
- Discuss the conceptual ideas behind various causal data science tools and algorithms
- Understand the importance of management theory for causal inference in business intelligence
- Carry out state-of-the-art causal data analyses by themselves

Examination

Exam ECTS	7.5
Examination form	Home assignment – written product
Individual or group exam	Individual exam
Assignment type	Report
Duration	2 weeks to prepare
Examiner	One internal examiner
Exam period	12/6/2021 – 12/20/2021
Aids	Open book: all written, electronic, and software aids
Make-up exam/re-exam	Same examination form as the ordinary exam

Course Content

Most managerial decision problems require answers to questions such as "*what happens if?*", "*what is the effect of X on Y*?", or "*was it X that caused Y to go up?*" In other words, practical business decision-making requires knowledge about cause-and-effect. While standard tools in machine learning and AI are designed for efficient pattern detection in high-dimensional settings, they are not able to distinguish causal relationships from simple correlations in the data. That means, most commonly used approaches to machine learning fall short in addressing pressing questions in business analytics and strategic management. This creates an important mismatch between the answers that these algorithms can provide and the problems that managers and strategists would like to solve. Which is why, in recent years, several leading companies from the tech sector and elsewhere (among them: Facebook, Google, Uber, Spotify, Zalando and McKinsey) have started to heavily invest into their causal data science capabilities.

This course will provide an introduction into the topic of causal inference in machine learning and AI, with a focus on applications to practically relevant, data-driven business cases. The course is meant to be conceptual rather than technical, in order to bridge the gap between data science and management strategy, for better evidence-based decision-making. A variety of hands-on examples will be discussed that allow students to apply their newly obtained knowledge and carry out state-of-the-art causal analyses by themselves. The course will thereby loosely follow the structure of *"The Book of Why"* by Judea Pearl and Dana Mackenzie, which has ushered a new era of causal thinking in data science and machine learning upon its publication in 2018. In particular, students will be put into the position to detect sources of confounding influence factors that threaten valid causal conclusions, understand the problem of selective measurement in data collection, and extrapolate causal knowledge across different business contexts. By developing an overarching framework for causal data science, the course will also cover several standard tools for causal inference, which are often used in empirical research in business and economics (such as difference-in-differences, instrumental variables, regression discontinuity designs, A/B testing and experiments, etc.). Thus, while not a research methods course as such, this elective will nonetheless be highly relevant for students who plan to conduct a quantitative data analysis as part of their master thesis project.

Teaching methods

The course consists of in-class lectures, guest lectures by practitioners from the tech sector, and hands-on tutorials in which students will learn how to carry out their own causal data analyses. In these practical sessions, state-of-the-art software for causal analysis will be used (www.causalfusion.net, no coding experience required). The course will incorporate (non-graded) problem sets, which can be done either individually or in groups, and which will prepare students for the written take-home exam. No specific prior knowledge is required. However, basic concepts in statistics (conditional means, variances, hypothesis testing, regression) will be useful and therefore repeated at the beginning of the course. In-class lectures will feature case studies and guest speakers to demonstrate the practical relevance of the covered material.

Course coordinator

Dr. Paul Hünermund
Copenhagen Business School
Department of Strategy and Innovation
Email: phu.si@cbs.dk

Office hours

Office hours are by appointment. Please contact the course coordinator via email.

Course Program

Week 35	**Introduction to the Course**
Topics:	• Causal knowledge for business decision-making • Ladder of causation
Required reading:	• **Syllabus** • Introduction – Pearl, J., and D. Mackenzie (2018). *The Book of Why*. Basic Books, New York.
Additional references:	• Google controversy, NYT: https://www.nytimes.com/2019/03/04/technology/google-gender- pay- gap.html • Morris, J. *Israeli data: How can efficacy vs. severe disease be strong when 60% of hospitalized are vaccinated?* https://www.covid-datascience.com/post/israeli-data-how-can-efficacy-vs-severe-disease-be-strong-when-60-of-hospitalized-are-vaccinated • Chapter 1 – Pearl, J., M. Glymour, and N. Jewell (2016). *Causal Inference in Statistics – A Primer*. Wiley: New Jersey, USA.

Week 36	**Graphical Causal Models I**
Topics:	• Directed acyclic graphs • D-separation & testable implications • Interventions in structural causal models • Backdoor criterion
Required reading:	• Chapter 1 – The Book of Why • Sections 1 & 2 in Hünermund, P., and E. Bareinboim (2021). Causal Inference and Data Fusion in Econometrics. https://arxiv.org/abs/1912.09104
Additional references:	• Chapter 2 – Pearl, J., M. Glymour, and N. Jewell (2016). *Causal Inference in Statistics – A Primer*. Wiley: New Jersey, USA.

Week 37	**Software Exercise** (online, self-paced)
Topics:	• Introduction to causalfusion.net
Required reading:	• Chapter 2 – The Book of Why
Additional references:	-

Week 38	**Graphical Causal Models II**
Topics:	• Matching and Regression • Front-door criterion • Causal discovery
Required reading:	• Chapter 3 – The Book of Why • Sections 3.1–3.2 in Hünermund & Bareinboim (2021)
Additional references:	• Chapter 3 – Pearl, J., M. Glymour, and N. Jewell (2016). *Causal Inference in Statistics – A Primer*. Wiley: New Jersey, USA. • Pearl, J. (1995). Causal diagrams for empirical research. *Biometrika*, 82(4), 669–710 • Neal, B. *The PC Algorithm for Causal Discovery*. https://www.youtube.com/watch?v=o2A61bJ0UCw

Week 40	**Experiments**
Topics:	• Randomized control trials • A/B testing in business • Difference-in-differences
Required reading:	• Chapter 4 – The Book of Why

	• Kohavi, R., and S. Thomke (2017). The Surprising Power of Online Experiments. *Harvard Business Review*. https://hbr.org/2017/09/the-surprising-power-of-online-experiments • Bojinov, I., G. Saint-Jacques, and M. Tingley (2020). Avoid the Pitfalls of A/B Testing. *Harvard Business Review*. https://hbr.org/2020/03/avoid-the-pitfalls-of-a-b-testing
Additional references:	• Chapter 9 – Cunningham, S. (2021). *Causal Inference: The Mixtape*. Yale University Press: New Haven, USA. Available here: https://mixtape.scunning.com

Week 41	**Surrogate Experiments**
Topics:	• *Z*-identification • Instrumental variables • Regression discontinuity designs
Required reading:	• Chapter 5 – The Book of Why • Section 3.4 in Hünermund & Bareinboim (2021) • Flammer, C. and P. Bansal (2017). Does a long-term orientation create value? Evidence from a regression discontinuity. *Strategic Management Journal*, 38(9): 1827–47
Additional references:	• Chapters 6 & 7 – Cunningham, S. (2021). *Causal Inference: The Mixtape*. Yale University Press: New Haven, USA. Available here: https://mixtape.scunning.com

Week 42	**Application of Causal Inference in Business** (online, self-paced)
Topics:	• *"In defense of curve fitting: how experimentation-driven and ML-enabled causal inference drives impact at Lyft"* – Keynote address at CDSM20 by Sean J. Taylor (Rideshare Labs, Lyft) https://causalscience.netlify.app/programme/keynote-videos
Required reading:	• Chapter 6 – The Book of Why • Hünermund, P., J. Kaminski, and C. Schmitt (2021). Causal Machine Learning and Business Decision Making. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3867326
Additional references:	• Doupe, P (Zalando SE). *How to Push Causal Inference in Industry?* https://causalscience.org/blog/how-to-push-causal-inference-in-industry

Week 44	**Causal Artificial Intelligence**
Topics:	• Do-calculus • Identification algorithms • Data fusion paradigm
Required reading:	• Chapter 7 – The Book of Why • Section 3.3 in Hünermund & Bareinboim (2021) • Hünermund, P. *What is Causal Data Fusion?* https://causalscience.org/blog/what-is-causal-data-fusion
Additional references:	• Salazar, D. Causality: Testing Identifiability. https://david-salazar.github.io/2020/07/31/causality-testing-identifiability/

Week 45	**Sample Selection Bias**
Topics:	• Selection diagrams • Recovering from selection bias in causal diagrams • Selection propensity score • Heckman selection model
Required reading:	• Chapter 8 – The Book of Why • Section 4 in Hünermund & Bareinboim (2021)
Additional references:	• Knox, D., W. Lowe, and J. Mummolo (2020). Administrative Records Mask Racially Biased Policing. *American Political Science Review,* 114(3): 619–637 • Angrist, J. (1997). Conditional independence in sample selection models. *Economics Letters*, 54, 103–112. • Heckman, J. (1979). Sample Selection Bias as a Specification Error. *Econometrica*, 47, 153–161

Week 47	**Counterfactuals**
Topics:	• Potential outcomes framework • Ignorability • Mediation and causal mechanisms • Fairness in algorithmic decision-making
Required reading:	• Chapter 9 – The Book of Why • Satell, G. and Y. Abdel-Magied (2020). AI Fairness Isn't Just an Ethical Issue. *Harvard Business Review*. https://hbr.org/2020/10/ai-fairness-isnt-just-an-ethical-issue
Additional references:	• Holland, P. (1986). Statistics and Causal Inference. *Journal of the American Statistical Association*. 81: 945–960. • Chapter 4 – Pearl, J., M. Glymour, and N. Jewell (2016). *Causal Inference in Statistics – A Primer*. Wiley: New Jersey, USA.

	• Zhang, J. and E. Bareinboim (2017). Fairness in Decision-Making – The Causal Explanation Formula. *Technical Report R-30-L*. Causal AI Lab, Purdue University. https://causalai.net/r30.pdf

Week 48	**External Validity**
Topics:	• External validity • The transportability problem • Meta-transportability
Required reading:	• Chapter 10 – The Book of Why • Section 5 in Hünermund & Bareinboim (2021)
Additional references:	• Townsend, J. (Microsoft). A/B Testing and Covid-19: Data-Driven Decisions in Times of Uncertainty. https://www.microsoft.com/en-us/research/group/experimentation-platform-exp/articles/a-b-testing-and-covid-19-data-driven-decisions-in-times-of-uncertainty/