

# SO9D6-20 Foundations of Social Data Science

**26/27**

**Department**

Sociology

**Level**

Taught Postgraduate Level

**Module leader**

Ulf Liebe

**Credit value**

20

**Module duration**

10 weeks

**Assessment**

100% coursework

**Study location**

University of Warwick main campus, Coventry

---

## Description

### Introductory description

What is so exciting about big data, data science and R? How can data science help towards a better understanding of social, political and economic processes? What does machine learning have to do with explaining societal outcomes, human behaviour and language usage? This module will develop students' understanding of social data science, corresponding ethical issues as well as why data science is a useful tool for research conducted at universities, governmental and non-governmental organisations, as well as private companies in various sectors. Besides discussing applications of social data science to the study of sociological, political, economic, criminal and language phenomena, the module will provide students with the skills to conduct and critically reflect on social data science research. The module does not require prior knowledge of R and advanced statistical techniques. It will provide an introduction to R and include some basic applications of social data science, for example in the context of web data collection and analysis, as well as quantitative text analysis.

### Module aims

This module will develop students' understanding of social data science, corresponding ethical issues as well as why data science is a useful tool for research conducted at universities,

governmental and non-governmental organisations, as well as private companies in various sectors. Besides discussing applications of social data science to the study of sociological, political, economic, criminal and language phenomena, the module will provide students with the skills to conduct and critically reflect on social data science research.

## Outline syllabus

This is an indicative module outline only to give an indication of the sort of topics that may be covered. Actual sessions held may differ.

Week 1: Understanding the Excitement: What is Social Data Science?

- This session will introduce social data science and refer to various applications across social science disciplines and the humanities.

Week 2: Exploring Novel Insights: What Does Social Data Science Tell Us About Societies?

- This session will provide a more in-depth discussion of selected applications of social data science as well as related ethical problems.

Week 3: Learning a Language of Data Science: Why Do Many People Talk about R, and How Does It Work?

- This session will provide an introduction to R. Also, access to (big) open data as well as limitations in data access of companies will be discussed.

Week 4: Harvesting Big Data: How Can We Collect Data from the Web? (I)

- Students will discuss applications of web data collection and analysis in the social sciences and humanities. They will apply the method to data such as Wikipedia pages.

Week 5: Harvesting Big Data: How Can We Collect Data from the Web? (II)

- Students will continue to apply the method to data such as (social) media content.

Week 6: Reading Week

Week 7: Predicting Social Outcomes: How Does Supervised Machine Learning Work? (I)

- Students will discuss applications of supervised machine learning in the social sciences and humanities. They will apply the method, for example to quantitative text data.

Week 8: Predicting Social Outcomes: How Does Supervised Machine Learning Work? (II)

- Students will continue to apply supervised machine learning.

Week 9: Discovering Social Structures: How Does Unsupervised Machine Learning Work? (I)

- Students will discuss applications of unsupervised machine learning in the social sciences and humanities. They will apply the method, for example to quantitative text data.

Week 10: Discovering Social Structures: How Does Unsupervised Machine Learning Work? (II)

- Students will continue to apply unsupervised machine learning.

### Illustrative Bibliography

Bisazza, A. and M. Federico, 2016. A survey of word reordering in statistical machine translation: Computational models and language phenomena. *Computational Linguistics*, 42(2), pp. 163-205.

Boumans, J.W. and D. Trilling, 2016. Taking stock of the toolkit. An overview of relevant automated content analysis approaches and techniques for digital journalism scholars. *Digital Journalism*, 1, 8-23.

Blumenstock, J., G. Cadamuro G and R. On, 2015. Predicting poverty and wealth from mobile phone metadata. *Science*, 350, pp.1073-1076.

Burscher, B., D. Odijk, R. Vliegthart, M. de Rijke and C.H. de Vreese, 2014. Teaching the computer to code frames in news: Comparing two supervised machine learning approaches to frame analysis. *Communication Methods and Measures*, 8 (3), pp. 190-206.

Cederman L.E. and N.B. Weidmann, 2017. Predicting armed conflict: Time to adjust our expectations? *Science*, 355, pp. 474-476.

Cranmer S.J. and B.A. Desmarais, 2017. What Can We Learn from Predictive Modeling? *Political Analysis*, 25, 145-166.

Connelly, R., C.J. Playford, V. Gayle and C. Dibben, 2016. The role of administrative data in the big data revolution in social science research. *Social Science Research*, 59, pp. 1-12

DiMaggio P., M. Nag and D. Blei, 2013. Exploiting affinities between topic modeling and the sociological perspective on culture: Application to newspaper coverage of U.S. government arts funding. *Poetics*, 41, pp. 570-606.

Evans J.A. and P. Aceves, 2016. Machine Translation: Mining Text for Social Theory. *Annual Review of Sociology*, 42, pp. 21-50.

Hastie T., R. Tibshirani and J. Friedman, 2009. *The Elements of Statistical Learning. Data Mining, Inference, and Prediction*. New York: Springer.

Healy, K. and J. Moody, 2014. Data Visualization in Sociology. *Annual Review of Sociology*, 40, 105-28.

Hofman, J.M., A. Sharma and D.J. Watts, 2017. Prediction and explanation in social systems. *Science*, 355, pp. 486-488.

Hopkins, D.J. and G. King, 2010. A Method of Automated Nonparametric Content Analysis for Social Science. *American Journal of Political Science*, 54, pp. 229-247.

Jordan, M.I. and T.M. Mitchell, 2015. Machine learning: Trends, perspectives, and prospects. *Science*, 349, pp. 255-260.

Kozlowski, A.C., M. Taddy and J.A. Evans, 2019. The Geometry of Culture: Analyzing the Meanings of Class through Word Embeddings. *American Sociological Review*, 84(5), pp. 905-949.

Lazer, D. and J. Radform, 2017. Data Ex Machina: Introduction to Big Data. *Annual Review of*

Sociology, 43, pp. 19-39.

Li, M., N. Turki, C.R. Izaguirre, C. DeMahy, B. Labery and T.T. Gage, 2020. Twitter as a tool for social movement: An analysis of feminist activism on social media communities. *Journal of Community Psychology* (online first).

McClendon, L. and N. Meghanathan, 2015. Using Machine Learning Algorithms to Analyse Crime Data. *Machine Learning and Applications: An International Journal (MLAIJ)*, 2(1), 1-12.

McFarland D.A., D. Ramage, J. Chuang, J. Heer, C.D. Manning, and D. Jurafsky, 2013. Differentiating language usage through topic models. *Poetics*, 41, pp. 607-625.

Mohr, J.W. and P. Bogdanov, 2013. Introduction Topic models: What they are and why they matter. *Poetics*, 41, pp. 545-569.

Pearl, J. and D. Mackenzie, 2018. *The book of why: The new science of cause and effect*. New York: Basic Books.

Salganik, M.J., 2017. *Bit by Bit: Social Research in the Digital Age*. Princeton & Oxford: Princeton University Press.

Sloan, L. and A. Quan-Haase, 2017. *Sage Handbook of Social Media Research Methods*. London: Sage.

Welbers, K., W. Van Atteveldt and K. Benoit, 2017. Text analysis in R. *Communication Methods and Measures*, 11(4), pp. 245-265.

Williams, M.L., P. Burnap and L. Sloan, 2017. Towards an ethical framework for publishing Twitter data in social research: taking into account users' views, online context and algorithmic estimation. *Sociology*, 51(6), pp. 1149-1168.

Williams, M.L., P. Burnap, H. Liu, A. Javed and A. Ozalp, 2020. Hate in the machine: Anti-black and anti-Muslim social media posts as predictors of offline racially and religiously aggravated crime. *British Journal of Criminology*, 60(1), pp. 93-117.

## **Learning outcomes**

By the end of the module, students should be able to:

- To familiarise students with foundations and applications of social data science
- To raise students' awareness of the potential and pitfalls of social data science
- To equip students with the skills to understand and undertake social data science research

## **Research element**

Some data analysis required in teaching

## **Interdisciplinary**

Interdisciplinary module for students in the social sciences and humanities

Opportunities for interdisciplinary learning are communicated to students

## Subject specific skills

Systematic understanding, coherent and detailed knowledge of key concepts and approaches of social data science

Ability to describe and comment on the advantages and pitfalls of different data science approaches

Ability to conduct social data science research

## Transferable skills

By developing and conducting their own data analysis ...  
the exercise of initiative and personal responsibility,  
decision-making in complex and unpredictable contexts when working with new data.

---

## Study

### Study time

Type	Required
Lectures	9 sessions of 1 hour (4%)
Seminars	9 sessions of 2 hours (9%)
Private study	173 hours (86%)
Total	200 hours

### Private study description

Reading for seminars; preparation for seminars; preparation for data analyses; preparation and writing of formative work; preparation and writing of summative work

### Costs

No further costs have been identified for this module.

---

## Assessment

You must pass all assessment components to pass the module.

Students can register for this module without taking any assessment.

## Assessment group A1

	Weighting	Study time	Eligible for self-certification
<b>Assessment component</b>			
Assessed essay Individual 4000 words essay	100%		Yes (extension)

Reassessment component is the same

## Feedback on assessment

Regular informal feedback will be provided throughout the module seminar sessions.

Formative: Feedback will be provided on the formative essay.

Summative: Written feedback will be provided on the summative essay.

---

## Availability

### Courses

This module is Optional for:

- All social science and humanities courses