Pensum STV2022

Solveig Bjørkholt[*]   Martin Søyland[†]

06-04-2022

Generelt

Under er en liste over bidrag som vil bli dekket kontinuerlig gjennom hele kurset:

1. Silge and Robinson (2017)
2. Grimmer, Roberts, and Stewart (2022)
3. Benoit et al. (2017)
4. Jurafsky and Martin (2021) (anbefalt)
5. Wickham (2016) (anbefalt)

Uke 34. Introduksjon (Solveig og Martin)

1. Grimmer, Roberts, and Stewart (2022) kap. 1-2 og 22 (36 sider)
2. Lucas et al. (2015) (24 sider)
3. Silge and Robinson (2017) kap. 1 (9 sider)
4. Pang, Lee, and others (2008) kap. 1 (10 sider)

Uke 35. Anskaffelse og innlasting av tekst (Martin)

1. Grimmer, Roberts, and Stewart (2022) kap. 3-4 (14 sider)
2. Cooksey (2014) kap. 1 (4 sider)
3. Wickham (2020) (8 sider)
4. Høyland and Søyland (2019) (22 sider)

Uke 36. Forbehandling av tekst 1 (Martin)

1. Grimmer, Roberts, and Stewart (2022) kap. 5 (11 sider)
2. Silge and Robinson (2017) kap. 3 (9 sider)
3. Jørgensen et al. (2019) (10 sider)
4. Barnes et al. (2019) (12 sider)
5. Benoit and Matsuo (2020) (11 sider)

Seminar 1: Anskaffe tekst og lage dtm i R

Uke 37. Forbehandling av tekst 2 (Solveig)

1. Grimmer, Roberts, and Stewart (2022) kap. 9 (7 sider)
2. Silge and Robinson (2017) kap. 4 (15 sider)

---
[*]solveig.bjorkholt@stv.uio.no
[†]martin.soyland@stv.uio.no

3. Denny and Spirling (2018) (21 sider)

Uke 38. Bruke API – Case: Stortinget (Martin)

1. *Stortingets Datatjeneste* (2022) (1 sider)
2. Søyland (2022) (1 sider)
3. Finseraas, Høyland, and Søyland (2021) (10 sider)

Seminar 2: Preprosessering av tekstdata i R

Uke 39. INGEN UNDERVISNING

Uke 40. INGEN UNDERVISNING

Uke 41. Veiledet læring og ikke-veiledet læring (Solveig)

1. Grimmer, Roberts, and Stewart (2022) kap. 10 og 17 (24 sider)
2. D'Orazio et al. (2014) (18 sider)
3. Feldman and Sanger (2006a) (17 sider)
4. Feldman and Sanger (2006b) (11 sider)
5. Muchlinski et al. (2016) (16 sider)

Uke 42. Ordbøker, tekstlikhet og sentiment (Solveig)

1. Grimmer, Roberts, and Stewart (2022) kap. 7 og 16 (12 sider)
2. Silge and Robinson (2017) kap. 2 (11 sider)
3. Pang, Lee, and others (2008) kap. 3-4 (26 sider)
4. Liu (2015a) (15 sider)
5. Liu (2015b) (30 sider)

Seminar 3: Sup vs. unsup i R

Uke 43. Klassifisering av tekst – Temamodellering (Martin)

1. Grimmer, Roberts, and Stewart (2022) kap. 13 og (13 sider)
2. Blei (2012) (8 sider)
3. Silge and Robinson (2017) kap. 6 (14 sider)
4. Roberts et al. (2014) (19 sider)

Uke 44. Estimere latent posisjon fra tekst (Solveig)

1. Laver, Benoit, and Garry (2003) (20 sider)
2. Slapin and Proksch (2008) (18 sider)
3. Lowe (2017) (15 sider)
4. Lauderdale and Herzog (2016) (20 sider)
5. Peterson and Spirling (2018) (8 sider)

Seminar 4: Modelleringsmetoder i R

Uke 46. Oppsummering (Solveig og Martin)

1. Grimmer, Roberts, and Stewart (2022) kap 28 (5 sider)

2. Wilkerson and Casas (2017) (19 sider)

Seminar 5: Fra tekst til funn – Q&A/Oppg.hjelp

Antall sider

[1] 574

Referanser

Barnes, Jeremy, Samia Touileb, Lilja Øvrelid, and Erik Velldal. 2019. "Lexicon Information in Neural Sentiment Analysis: A Multi-Task Learning Approach." In *Proceedings of the 22nd Nordic Conference on Computational Linguistics*, 175–86. Turku, Finland: Linköping University Electronic Press. https://aclanthology.org/W19-6119.

Benoit, Kenneth, and Akitaka Matsuo. 2020. *Spacyr: Wrapper to the 'spaCy' 'NLP' Library*. https://CRAN.R-project.org/package=spacyr.

Benoit, Kenneth, Kohei Watanabe, Paul Nulty, Adam Obeng, Haiyan Wang, Benjamin Lauderdale, and Will Lowe. 2017. *Quanteda: Quantitative Analysis of Textual Data*. http://quanteda.io.

Blei, David M. 2012. "Probabilistic Topic Models." *Communications of the ACM* 55 (4): 77–84.

Cooksey, Brian. 2014. "An Introduction to APIs." *Zapier, Inc. Cvetojevic, S., Juhasz, L., & Hochmair, H.(2016). Positional Accuracy of Twitter and Instagram Images in Urban Environment. Https://Doi. Org/10.1553/Giscience2016_01_s191*.

D'Orazio, Vito, Steven T. Landis, Glenn Palmer, and Philip Schrodt. 2014. "Separating the Wheat from the Chaff: Applications of Automated Document Classification Using Support Vector Machines." *Political Analysis* 22 (2): 224–42. https://doi.org/https://doi.org/10.1093/pan/mpt030.

Denny, Matthew J., and Arthur Spirling. 2018. "Text Preprocessing For Unsupervised Learning: Why It Matters, When It Misleads, And What To Do About It." *Political Analysis* 26 (2): 168–89.

Feldman, Ronen, and James Sanger. 2006a. "Categorization." In *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*, 64–81. Cambridge University Press. https://doi.org/10.1017/CBO9780511546914.005.

———. 2006b. "Clustering." In *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*, 82–93. Cambridge University Press. https://doi.org/10.1017/CBO9780511546914.006.

Finseraas, Henning, Bjørn Høyland, and Martin G. Søyland. 2021. "Climate Politics in Hard Times: How Local Economic Shocks Influence MPs Attention to Climate Change." *European Journal of Political Research* 60 (3): 738–47. https://doi.org/https://doi.org/10.1111/1475-6765.12415.

Grimmer, Justin, Margaret E. Roberts, and Brandon M. Stewart. 2022. *Text as Data: A New Framework for Machine Learning and the Social Sciences*. Princeton University Press.

Høyland, Bjørn, and Martin Søyland. 2019. "Electoral Reform and Parliamentary Debates." *Legislative Studies Quarterly* 44 (4): 593–615. https://doi.org/10.1111/lsq.12237.

Jurafsky, Daniel, and James H. Martin. 2021. *Speech and Language Processing*. Online draft.

Jørgensen, Fredrik, Tobias Aasmoe, Anne-Stine Ruud Husevåg, Lilja Øvrelid, and Erik Velldal. 2019. "NorNE: Annotating Named Entities for Norwegian." arXiv. https://doi.org/10.48550/ARXIV.1911.12146.

Lauderdale, Benjamin E., and Alexander Herzog. 2016. "Measuring Political Positions from Legislative Speech." *Political Analysis* 24 (3): 374–94. https://doi.org/10.1093/pan/mpw017.

Laver, Michael, Kenneth Benoit, and John Garry. 2003. "Extracting Policy Positions from Political Texts Using Words as Data." *American Political Science Review* 97 (02): 311–31.

Liu, Bing. 2015a. "Introduction." In *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions*, 1–15. Cambridge University Press. https://www.cambridge.org/core/books/sentiment-analysis/3F0F24BE12E66764ACE8F179BCDA42E9.

———. 2015b. "The Problem of Sentiment Analysis." In *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions*, 16–46. Cambridge University Press. https://www.cambridge.org/core/services/aop-cambridge-core/content/view/A0AFE2C49D72C5914C34DAC763BCD931/9781139084789c2_p16-46_CBO.pdf/problem_of_sentiment_analysis.pdf.

Lowe, Will. 2017. "Understanding Wordscores." *Political Analysis* 16 (4): 356–71. https://doi.org/10.1093/pan/mpn004.

Lucas, Christopher, Richard A. Nielsen, Margaret E. Roberts, Brandon M. Stewart, Alex Storer, and Dustin Tingley. 2015. "Computer-Assisted Text Analysis for Comparative Politics." *Political Analysis* 23 (2): 254–77. https://doi.org/10.1093/pan/mpu019.

Muchlinski, David, David Siroky, Jingrui He, and Matthew Kocher. 2016. "Comparing Random Forest with Logistic Regression for Predicting Class-Imbalanced Civil War Onset Data." *Political Analysis* 24 (1): 87–103. https://doi.org/10.1093/pan/mpv024.

Pang, Bo, Lillian Lee, and others. 2008. "Opinion Mining and Sentiment Analysis." *Foundations and Trends in Information Retrieval* 2 (1–2): 1–135.

Peterson, Andrew, and Arthur Spirling. 2018. "Classification Accuracy as a Substantive Quantity of Interest: Measuring Polarization in Westminster Systems." *Political Analysis* 26 (1).

Roberts, Margaret E., Brandon M. Stewart, Dustin Tingley, Christopher Lucas, Jetson Leder-Luis, Shana Kushner Gadarian, Bethany Albertson, and David G. Rand. 2014. "Structural Topic Models for Open-Ended Survey Responses." *American Journal of Political Science* 58 (4): 1064–82.

Silge, Julia, and David Robinson. 2017. *Text Mining with R: A Tidy Approach.* O'Reilly Media, Inc.

Slapin, Jonathan B., and Sven-Oliver Proksch. 2008. "A Scaling Model for Estimating Time-Series Party Positions from Texts." *American Journal of Political Science* 52 (3): 705–22.

Stortingets Datatjeneste. 2022. https://data.stortinget.no.

Søyland, Martin. 2022. *Stortingscrape: Scrape and Structure Raw Data from the Norwegian Parliament's API.* https://github.com/martigso/stortingscrape.

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis.* Springer.

———. 2020. *Httr: Tools for Working with URLs and HTTP.* https://cran.r-project.org/web/packages/httr/vignettes/quickstart.html.

Wilkerson, John, and Andreu Casas. 2017. "Large-Scale Computerized Text Analysis in Political Science: Opportunities and Challenges." *Annual Review of Political Science* 20 (1): 529–44. https://doi.org/10.1146/annurev-polisci-052615-025542.