Pensum STV2022

Eli Sofie Baltzersen* Martin Søyland†

19-04-2023

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 (Eli og Martin)

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

Uke 35. Anskaffelse og innlasting av tekst (Martin)

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

Uke 36. Forbehandling av tekst 1 (Martin)

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

Seminar 1: Anskaffe tekst og lage dtm i R

Uke 37. Forbehandling av tekst 2 (Eli)

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

*e.s.baltzersen@stv.uio.no
†martin.soyland@stv.uio.no

3. Denny and Spirling (2018)

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

1. Stortinget (2022)
2. Søyland (2022)
3. Finseraas, Høyland, and Søyland (2021)

Seminar 2: Preprosessering av tekstdata i R

Uke 39. INGEN UNDERVISNING

Uke 40. INGEN UNDERVISNING

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

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

Uke 42. AI, maskinlæring og dyp læring (Martin)

1. Grimmer, Roberts, and Stewart (2022) kap 19
2. Chatsiou and Mikhaylov (2020)
3. Olivella and Shoub (2020)
4. Ongsulee (2017)

Uke 43. Ordbøker, tekstlikhet og sentiment (Eli)

1. Grimmer, Roberts, and Stewart (2022) kap. 7 og 16
2. Silge and Robinson (2017) kap. 2
3. Pang, Lee, et al. (2008) kap. 3-4
4. Liu (2015a)
5. Liu (2015b)

Seminar 3: Sup vs. unsup i R

Uke 44. Klassifisering av tekst – Temamodellering (Martin)

1. Grimmer, Roberts, and Stewart (2022) kap. 13 og

2. Blei (2012)
3. Silge and Robinson (2017) kap. 6
4. Roberts et al. (2014)

Seminar 4: Modelleringsmetoder i R

Uke 46. Oppsummering (Eli og Martin)

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

2. Wilkerson and Casas (2017)

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

Antall sider

```
## [1] 0
```

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.

Chatsiou, Kakia, and Slava Jankin Mikhaylov. 2020. "Deep Learning for Political Science." In. https://doi.org/10.4135/9781526486387.

Cooksey, Brian. 2014. "An Introduction to APIs." *Zapier, Inc.* https://cdn.zapier.com/storage/learn_ebooks/e06a35cfcf092ec6dd22670383d9fd12.pdf.

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. https://doi.org/10.1017/pan.2017.44.

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://ejpr.onlinelibrary.wiley.com/doi/abs/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.

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

Jurafsky, Daniel, and James H. Martin. 2021. *Speech and Language Processing*. Online draft. https://web.stanford.edu/~jurafsky/slp3/.

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.

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.

Olivella, Santiago, and Kelsey Shoub. 2020. "Machine Learning in Political Science: Supervised Learning Models." In. https://doi.org/10.4135/9781526486387.n59.

Ongsulee, Pariwat. 2017. "Artificial Intelligence, Machine Learning and Deep Learning." In *2017 15th International Conference on ICT and Knowledge Engineering (ICT&KE)*, 1–6. https://doi.org/10.1109/ICTKE.2017.8259629.

Pang, Bo, Lillian Lee, et al. 2008. "Opinion Mining and Sentiment Analysis." *Foundations and Trends® in Information Retrieval* 2 (1–2): 1–135. https://www.cs.cornell.edu/home/llee/omsa/omsa.pdf.

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. https://www.tidytextmining.com/.

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

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

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer. https://ggplot2-book.org/.

———. 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.