

STV2022

Store tekstdata:

Samling, strukturering og analyse av politiske tekster

Martin Søyland* og Eli S. Baltzersen†

November 23, 2023

Kort om emnet

Store mengder tekst produseres og publiseres på internett hver dag. Mange av disse tekstene er relevante for statsvitenskapelige spørsmål, som taler på Stortinget, lovforslag fra regjeringen, politiske debattinnlegg i aviser eller diskusjoner i sosiale medier. Mer tilgjengelige digitaliserte politiske tekster i det offentlige har åpnet et mulighetsrom for tekstanalyse som ikke var mulig eller var veldig vanskelig og tidkrevende før. Samtidig utvikles stadig nye metoder innenfor maskinlæring, som muliggjør analyse av disse store tekstdataene.

Det kan ofte være vanskelig å finne mønster som kan svare på spørsmål og teorier vi har i statsvitenskap i disse store tekstsamlingene. Derfor kan vi se til metoder innenfor maskinlæring for å analysere store samlinger av tekst systematisk. Samtidig er det ikke alltid digitaliserte politiske tekster er tilrettelagt for å analyseres direkte. I disse tilfellene er god strukturering av rådata viktig.

I dette kurset vil du lære om hele prosessen for hvordan man kan benytte seg

*E-mail: martin.g.soyland@gmail.com

†E-mail: e.s.baltzersen@stv.uio.no

av slike kvantitative analysemetoder på tekstdata for å svare på statsvitenskapelige spørsmål: Hvordan samler man effektivt og redelig store mengder politiske tekster? Hva må til for å gjøre slike tekster klare for analyse? Og hvordan kan vi analysere tekstene?

Hva lærer du?

Kunnskaper

Du vil lære:

- teknikker for effektiv samling av store mengder tekst fra offentlige kilder
- strukturere og klargjøre dokumentsamlinger for kvantitativ analyse
- bruksområder for tekst i statsvitenskaplig analyse
- analysemetoder for å trekke ut meningsfulle innsikter fra dokumentsamlinger

Ferdigheter

Du vil kunne:

- samle store mengder data fra strukturerte og ustrukturerte offentlige kilder
- klargjøre store dokumentsamlinger for analyse
- identifisere passende analyseteknikk for å besvare en problemstilling
- oppsummere og presentere analyseresultater av tekst på en forståelig måte

Generell kompetanse

Du vil kunne:

- gjøre utvalg og tilpasse kvantitative tekstanalyseteknikker for eget arbeid
- samle, strukturere, preprosessere og endre store mengder tekst
- anvende forskjellige analysemetoder og oppsummere resultater på en meningsfull måte
- utføre arbeidslivsrelevant tekstbehandling og analyse

Undervisning

Forelesninger og seminarer.

Obligatoriske aktiviteter:

- Oppmøte på første seminar
- Oppmøte på to av fire påfølgende seminarer
- Lovere oppgave etter hvert seminar

Se SV-fakultetets regler for bytte av seminarer og retningslinjer for obligatoriske aktiviteter.

Fravær fra obligatorisk aktivitet:

Ved sykdom eller andre tungtveiende grunner kan du i noen tilfeller få godkjent gyldig fravær eller utsettelse av obligatorisk aktivitet.

Meld fra om fravær fra eller behov for utsettelse av obligatorisk aktivitet
Adgang til undervisning:

Dersom du har gjennomført og fått godkjent obligatorisk undervisning, har du ikke krav på ny undervisning. Dersom du har fått undervisningsopptak til emnet, men ikke har gjennomført eller fått godkjent obligatorisk undervisning, har du rett til ny undervisning når det er ledig kapasitet.

Forelesninger

Forelesningene vil dekke hovedtemaene i kurset med eksempler fra forskning og arbeidsliv. Tema for forelesning og seminar vil bli delt opp i hvordan å:

- anskaffe politiske tekster i forskjellige format
- strukturere tekster i form av kvantitative data
- oppsummere mengder tekst
- finne riktig analysemetode til tekstsamlinger for å svare på konkrete problemstillinger

Forslag til forelesninger:

1. Introduksjon
2. Anskaffelse og innlasting av tekst (a)
3. Bruke API – case: Stortinget
4. Forbehandling av tekst (a)
5. Forbehandling av tekst (b)
6. Veiledet læring (supervised learning) vs. ikke-veiledet læring (unsupervised learning)
7. Ordbøker og likhet mellom tekster
8. Klassifisering av tekst – temamodellering
9. Estimerer latent (ideologisk) posisjon fra tekst
10. Oppsummering

Seminarer

I seminarene vil studentene lære å utføre datasamling, strukturering av data og analyse av data, knyttet til forelesningene.

Seminarene vil fungere som et springbrett for deloppgavene over semesteret, der man går gjennom generelle trekk ved den aktuelle oppgaven.

Obligatoriske aktiviteter

- Oppmøte på seminarene
- Levere oppgave etter hvert seminar

Eksamen

Semesteroppgave med justerende muntlig eksamen.

Godkjente obligatoriske aktiviteter er et vilkår for å kunne avlegge eksamen.