

Til:	Programrådet i sosiologi
Fra:	Undervisningsleder i sosiologi
Sakstype:	Vedtaks sak
Møtedato:	08.12.2021
Notatdato:	01.12.2021
Saksbehandler:	Sara Fanny Maria Kvamme-Vik

SAK 6 (saksnr. 58/21) Opprettelse av nytt SOS-BA emne, SOS2901 – Anvendt Maskinlæring

Kurset gir en praktisk innføring i maskinlæring der man bruker data på tidligere hendelser til å gjøre klassifiseringer og prediksjoner for nye observasjoner der man ennå ikke vet utfallet.

Dette er metoder anvendes på mange felt og vi vektlegger praktiske anvendelser innenfor f.eks. risikovurderinger på justisfeltet, micro-targeting for politisk påvirkning, kundebehandling, sosiale medier, kredittvurderinger osv.

Prediksjoner vil aldri være perfekte, og vurdering av presisjon er sentralt tema. Herunder gjelder vurdering av ulike typer feil (falske positive vs. falske negative), og om det er systematiske skjevheter i presisjon for undergrupper (f.eks. kjønn, innvandringsbakgrunn, bosted). Det vektlegges hvordan vurdering av praktiske konsekvenser og rettferdighet har betydning for estimering og justering av algoritmene.

Det tekniske nivået er moderat og det vektlegges grunnleggende prinsipper og praktisk bruk av software. Undervisningen tar utgangspunkt i at studentene er kjent med grunnleggende regresjonsanalyse og lineære modeller. Undervisningen er praktisk rettet med bruk av softwaren *R* og *RStudio*.

Læringsmål

Etter endt kurs skal studentene kunne følgende:

- Kjenne viktige prinsipper for maskinlæring, herunder forstå bias-variance trade-off, overfitting osv. og teknikker for å håndtere disse.
- Grunnleggende ferdigheter i å bruke grunnteknikker for prediksjon og klassifisering, som regresjon, CART, bagging, random forest, boosting og supported vector machines. Herunder justere modellene med tuning-parametere.
- Vurdere algoritmenes pålitelighet, med vekt på feilrater og mål på algorithmic fairness
- Ha generell kunnskap om hvordan substansielle vurderinger har konsekvenser for estimering av modellene
- Forstå viktige muligheter og begrensninger ved datadrevne modeller. Herunder datakvalitet, forsterkning av bias over tid, og utilsiktede konsekvenser

Pensum

Spesifikt pensum blir lagt frem på pensummøtet september 2022, men her er et utkast:

Bruker en lærebok som dekker grunnleggende teknikker. Følgende bok er særlig aktuell fordi den er skrevet for studenter som er kjent med et standard regresjonsmodeller (som læres på SOSGEO1120).

Berk R (2016) *Statistical learning from a regression perspective*, Springer

En mer praktisk bok eller softwaremanual, f.eks. følgende:

<https://bradleyboehmke.github.io/HOML/>

<https://emilhvitefeldt.github.io/ISLR-tidymodels-labs/index.html>

Og en manual for algorithmic fairness:

<https://kozodoi.me/r/fairness/packages/2020/05/01/fairness-tutorial.html>

I tillegg en håndfull artikler som bruker anvendt litteratur som viser bruk av de ulike algoritmene med eksempler fra for eksempel følgende:

Risikovurderinger (løslatelse fra fengsel, predictive policing, tidlig intervensjon)

Sosiale medier og plattformer (Facebook, Twitter, Über)

Arbeidsliv (rekruttering, hindre turnover)

Offentlig sektor/ velferdssystemet (NAV, Skatteetaten)

Målrrette budskap (kundesegmentering, politisk påvirkning)

Undervisning

Undervisningen vil være basert på følgende komponenter:

- Forhåndsinnspilte korte forelesningsvideoer (20-30 minutter)
- Plenumssamlingene blir dialogbaserte / problemløsning
- Praktisk oppgaveløsning med R-programmering i web-basert verktøy (shiny-server på uio) med automatisk retting og tilbakemelding.

Forutsetninger

Det forutsettes en grunnleggende kjennskap til kvantitative metoder og statistikk. Studentene bør ha noe kjennskap til programmet R. Det forventede nivået tilsvarer fullført SOSGEO1120 eller STV1020. Studenter som ikke har tatt disse emnene bør sjekke læringsmålene på emnesiden til SOSGEO1120.

Studentene må ha tilgang til en datamaskin med internetttilgang.

Eksamen

4-timers skoleeksamen.

Eksamen vil være praktisk rettet med bruk av datasett og programmering i R.

Andre ting:

- Obligatorisk aktivitet: nei
- Undervisningsramme: tilsvarer 10 forelesninger på 2x45 minutter.
- Max antall studenter: 100
- Tilbys til hvilke studenter: åpent for alle, del av 40-gruppen i sosiologi. Del av 40-gruppen i datavitenskap (for ISS og ISV).
- Forutsetninger: nei. Men *anbefaling* om metodekurs.
- Andre krav: tilgang til egen pc med installert R og Rstudio
- Undervisningsspråk: norsk
- Undervises når og hvor ofte: vårsemester hvert år.

Vedtaksforslag: Programrådet vedtar opprettelsen av SOS2901 med oppstart våren 2023. Pensum og endelig emnebeskrivelse vedtas på pensummøtet høsten 2022.