

## Bakgrunn for endringer SOS2900

SOS2900 har fra oppstart hatt et fokus på maskinlæring med anvendelser innenfor justisfeltet, arbeidslivet og sosiale medier, samt overvåking og «big data». På den tekniske siden har både «supervised learning» og «unsupervised learning» blitt dekket. Dette har vært å favne litt vel bredt både tematisk og teknisk. Det er en fordel å fokusere det tematiske til noe mer avgrenset og da fokusere på det emneansvarlig kan best: justisfeltet. Teknisk fokuseres det mot «supervised learning», som i praksis er risikovurderinger og prediksjon om fremtiden eller om nye observasjoner.

Kurset har fremstått som teknisk og vanskelig i både navn og emnebeskrivelsen, noe som trolig har gitt lavere studenttall. Å fokusere temaet med tilhørende tittel vil forhåpentligvis gjøre kurset mer tiltalende for flere. Tittel og emnebeskrivelse er derfor endret.

Kurset grunnstruktur har mye likt med tidligere: hands-on metodisk kompetanse kombinert med substansiell litteratur og sosiologiske perspektiver. Det er stor overføringsverdi til anvendelser på andre felt.

Våren 2020 går ikke kurset SOS2300 (Avvik og marginalisering) fordi Willy har forskningsfri. Å fokusere SOS2900 mot kriminalitet kan dermed muligens være et aktuelt tilbud for noen av de som ellers kunne tenke seg å ta SOS2300.

Om det pedagogiske vil det bli viktige endringer:

- Vi kan ikke lengre bruke datacamp.com fordi det krever egen databehandlingsavtale med Datacamp, og det må vi stå for selv. Det prioriteres ikke.
- Det er under utvikling en e-læring system på SV-fakultetet med bruk av R med automatisk tilbakemelding. Arbeidet ledes av Bjørn Høyland (STV) og undertegnede er med i arbeidsgruppen. Dette skal implementeres våren 2020 og benyttes på SOS2900. Undervisningen legges derfor opp til kun forelesninger uten egen datalab. Dette gjør at kurset ikke lengre trenger begrensning på antall plasser.
- Det legges opp til to obligatoriske krav:
  - o Oppmøte på første forelesning. Dette har rent praktiske grunner: å vite hvor mange studenter som faktisk deltar. Det har betydning for hvordan undervisningen legges opp.
  - o Gruppeoppgave med «hverandrevurdering». De jobber gjennom hele semesteret med å analysere eget datasett parallelt med undervisningen. Dette munner ut i et essay som presenteres muntlig og får godkjent. Dette skal i utgangspunktet ikke kreve mer undervisningsressurser. (Hvis det blir veldig mange studenter på kurset kan det derimot noen praktiske utfordringer).

**Føringer for gruppeoppgave:** 3-4 studenter får skriver sammen og analyserer et datasettet med bruk av teknikker fra kurset og skal ha følgende innhold:

- En kort innledning: presentere data, utfall og hvorfor det er relevant med prediksjon/klassifikasjon for dette utfallet. Substansielt pensum skal benyttes i diskusjon.
- Presentere deskriptiv statistikk og grafikk
- Prediksjon/klassifikasjon med regresjon
- Prediksjon/klassifikasjon med maskinlæring. Feilrater, cost-ratio og bias skal diskuteres, og brukes til justering av modellen
- Alle deler skal kommenteres, og det skal skrives en konklusjon der det substansielle pensum benyttes

- Gruppeoppgavene skal leveres i Canvas før det presenteres i plenum (max 10 minutter).
- All kode skal vedlegges
- Gruppene bytter essay seg imellom og det skal gis skriftlige kommentarer til hverandres arbeid. Altså: studentene retter hverandre. Dette skal gjøres før muntlig fremlegging slik at innspill fra andre kan innarbeides før ferdigstilling.

Eksamen blir som i 2019: tradisjonell skoleeksamen der det inkluderes programmeringsoppgaver med bruk av R. Dette fungerte godt i fjor.